

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 20 (1995)
Indiana University

Recollection of Illusory Voices¹

**Helena M. Saldaña,² Kathleen B. McDermott,³ David B. Pisoni,
and Henry L. Roediger III³**

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹ This research was supported by NIH NIDCD Research Grant DC-00111 to Indiana University in Bloomington, IN.

² Now at House Ear Institute, Los Angeles, CA.

³ Psychology Department at Rice University, Houston TX.

Abstract

When subjects are given a list of words that are strongly associated with a nonpresented item they frequently recall the missing word and also recognize it as having occurred in the study list (Roediger & McDermott, 1995). Subjects also report that they remember the actual event of hearing the nonpresented word. The present investigation explores the phenomenology of false recognition in four experiments. In these studies, subjects were presented with auditory lists of related words spoken by male and female voices, and then given a recognition test containing both studied items and highly related nonpresented items. Subjects were asked to judge each test item as "old" or "new" and, if old, whether they remembered the presentation voice. In all four experiments, subjects showed high levels of false recognition and in general the same willingness to attribute the gender of the speaker to the nonpresented items as to the presented items. This pattern emerged for both lists with mixed voices (i.e., male and female) and lists with consistent voices (i.e., male or female); and for auditory as well as visual recognition tasks. The general pattern of results demonstrates that subjects' recognition of critical lures in this paradigm has the same essential characteristics of recognition of studied items, thereby producing a subjectively compelling memory illusion. In addition, the act of recall generally enhanced both veridical and illusory recognition.

Recollection of Illusory Voices

Roediger and McDermott (1995) presented subjects with lists of words that were strongly related to a nonpresented word and showed that subjects frequently recalled the nonpresented word on an immediate recall test. In subsequent recognition tests, subjects falsely recognized the nonpresented item at the same level (around 80%) as they did the studied items. In addition, when subjects were asked to make "remember/know" judgments on recognized items (Tulving, 1985), they claimed to *remember* the nonpresented items at about the same level as the presented items. Roediger and McDermott (1995) concluded that items falsely recalled and falsely recognized seemed indistinguishable to their subjects from other items/information recalled from secondary memory. These results seemed remarkable because subjects were recalling and recognizing word lists (believed, since Bartlett's (1932) pronouncements, to be recalled in a reproductive rather than a reconstructive manner), after short retention intervals and with specific instructions not to guess.

Roediger and McDermott's (1995) results have generally been confirmed by other researchers. Payne, Elie, Blackwell, and Neuschatz (1996) conducted several experiments that were rather close replications of Roediger and McDermott's (1995) Experiment 2, obtaining similar results (with one exception to be mentioned below) and extended the paradigm in several ways. McDermott (1996) showed that after a 2-day delay, false recall exceeded veridical recall; she also showed that the illusory false recall was still maintained, albeit at reduced levels, after five study/test cycles. That is, subjects still falsely recalled words after hearing the lists and being tested on it five times. Norman and Schacter (1996) and Schacter, Verfaellie, and Pradere (1996) used the Roediger-McDermott paradigm to test elderly subjects and amnesiac subjects, respectively. Interestingly, older subjects showed the illusory false recall and false recognition to a greater extent than did younger subjects, despite recalling the lists less well overall (as expected). On the other hand, amnesiac patients' recall of the lists items was very poor, and they failed to show the illusory recall and recognition effects.

These experiments and others (e.g., Read, 1996) generally all confirm the results of Roediger and McDermott (1995), even if leaving them poorly understood theoretically. The primary motivation for the present series of experiments was to examine subjects' phenomenological experience during recognition for both veridical and illusory recognition. In particular, we wished to examine subjects' attributions about the voice used for presentation when the lists of related words were presented in two voices. Previous research on recognition memory has demonstrated that listeners are quite good at remembering not only the specific words that were presented to them during study but also the physical attributes of the voice (Palmeri, Goldinger & Pisoni, 1993), suggesting that memory representations for spoken words contain highly specific details about the event of hearing the word (Goldinger, 1992). In this series of experiments, we were interested in whether the phenomenological experience in false recognition was similarly detailed. Another purpose of the present set of experiments was to gain insight into a more theoretical question about the nature of representations for illusory memories. There is now converging evidence that the phenomenological experience of an illusory memory is very similar to a veridical memory; if so, the suggestion could be made that the representations in memory are similar if not the same. In this set of experiments, we try to determine whether there are fundamental differences between the two types of representations that can be observed during recognition memory experiments.

While conducting this research, we became aware of related work by Payne et al. (1996), who asked a similar question. In their Experiment 3, subjects heard the Roediger-McDermott (1995) lists presented in two voices and then recalled the lists three times. After the third recall, subjects were asked to indicate beside each recalled item whether it had been presented in a male voice, a female voice, or to indicate that

they did not know which voice the item had originally been presented in. Under the conditions in Payne et al.'s experiment, the probability of both veridical and false recall was just over .30. Subjects made voice attributions for between 80 and 90% of the recalled items, both studied and nonstudied, and rarely used the "don't know" category even for items that had not been presented. In the experiments reported below, we asked subjects to assign input modality judgments during a recognition test, rather than after a recall test. Because false recognition in the Roediger-McDermott (1995) paradigm is often higher than 70%, subjects will have more opportunities to judge modality of occurrence for both studied and nonstudied (critical) items.

In all of our experiments, subjects heard lists of words presented in both a male and female voice and then later received a recognition test that requested information about whether test items had been previously heard in one voice or the other voice. Across four generally similar experiments we included manipulations that permitted us to ask the following questions: When experiencing a false recognition would subjects be as willing to attribute the critical lure being spoken in one voice or the other? Would such attributions be as frequent for illusory memories as for veridical memories? Answers to both questions were affirmative in Experiment 1, in which the lists were presented with some words in a male voice and other words in a female voice, and in which items on the recognition test were presented visually. In Experiment 2, we asked if the same outcome would be obtained if subjects received an auditory recognition test instead of a visual recognition test. Matching and mismatching of voice characteristics between study and test might provide an additional clue to recognition and could thereby reduce false recognition. In Experiments 3 and 4, we asked if presenting lists all in one voice or all in the other voice might bias a subject's response for voice attributions. In other words, are subjects more likely to attribute a male voice to a critical lure if the original associated list was presented in a male voice? We also asked whether pure voice lists would reduce false recognition when the test items were presented either visually or auditorily.

A secondary motivation for the current experiments was to confirm (or disconfirm) one aspect of the Roediger and McDermott (1995) results that has not been consistently replicated. They showed that prior recall of lists of words generally increased later recognition, both for veridical (studied) items and for illusory (nonstudied) items. In their Experiment 2 subjects heard 16 lists of words but recalled only half the lists. On the later recognition tests, both veridical and illusory recognition were greater for the previously recalled lists than for the nonrecalled lists. However, others have not consistently obtained these effects. Payne et al. (1996) found such a testing effect in recognition for studied items but not for nonstudied (critical) items. Schacter et al. (1996) did not find statistically significant testing effects for either studied or nonstudied items. Other experiments have reported generally small positive effects of a prior free recall test on later recognition for studied items, but only for items occurring at the end of the list (see Jones & Roediger, 1995; Lockhart, 1975). Therefore, the inconsistencies in the outcome may not be too surprising. However, because repeated testing seems to represent a key aspect to the development of false memories (see Roediger, McDermott & Goff, in press), the effect of recall on later recognition in this paradigm deserves more careful examination. McDermott (1996) has shown that prior recall of both studied and nonstudied items has a powerful effect on their later recall, so the only disputed question is whether prior recall affects later recognition. Therefore, each of the four experiments reported here incorporated the design of Roediger and McDermott's (1995) Experiment 2 in which subjects studied 16 lists and were tested on 8, so that the effect of recall on later recognition could be examined.

In sum, our experiments explored the issues of subjects' willingness to attribute recollection of voice information for both accurately recognized (studied) and false recognized (nonstudied) items and the effect of prior recall on both veridical and illusory recognition.

Experiment 1

The purpose of the first experiment was to replicate and extend the recent findings of Roediger & McDermott (1995). These researchers utilized a procedure developed by Tulving (1985), to try to gain some insight into the phenomenological experience of subjects when falsely recognizing critical lures (Experiment 2). The methodology involved asking listeners to make “remember/know” judgments on items that they had identified as old. A *remember* judgment was to be given if subjects could mentally relive the experience of the presentation of the item (recalling what they were doing when the item was presented, or the physical characteristics associated with the item presentation). A *know* judgment was to be given if the subject was confident that the item appeared on the list but could not remember anything specific about the actual occurrence of the item on the list.

The pattern of results showed that *remember* judgments were just as likely to be given to falsely remembered items as actual remembered items. It was concluded that false memories can be the result of conscious recollection and not only general familiarity. In Experiment 1 of the current investigation we extend this finding by asking subjects to make judgments on the source attributes of the list items. This methodology avoids asking subjects to report directly on their phenomenological experience. Instead, subjects are asked a specific question about the occurrence of a recognized item (whether it was presented in a male voice or female voice)

Method

Subjects

Thirty-two Indiana University undergraduates participated in a one-hour session in partial fulfillment of a course requirement in an introductory psychology course. All subjects were native speakers of English and reported no history of hearing or speech impairments at the time of testing.

Materials

The 24 word lists developed by Roediger and McDermott (1995) were utilized in the present experiment. Each list consisted of 15 words that were highly associated to a nonpresented word in the Russell and Jenkins (1954) norms. The resulting 360 items were digitized on-line at a 20-kHz sampling rate and 16-bit resolution by both a male and a female talker. The root-mean-square amplitude of all stimulus items was equated using a signal processing package. Each list was composed of alternating male and female voices; half of the lists started with a male voice and half of the lists started with a female voice. The ordering of each list was held constant with the highest associates occurring first.

Design

The 24 lists were divided into three sets for the purpose of counterbalancing. Subjects were presented with 16 lists during the study phase of the experiment, with 8 of the lists tested for immediate free recall and the other 8 lists not tested. The remaining 8 lists were not presented during the study phase but appeared on the subsequent recognition test. In the recognition test, subjects were presented with visual words and were asked to determine whether the item had been presented previously. If subjects judged that the item had not been presented during the study phase they were to press a button labeled “new.” If the subjects determined that an item was presented during the previous study phase, they were to give one of three possible responses. If the item had been presented previously and the subject remembered which voice it was presented in (male or female) then they were to press buttons labeled “male-old” or “female-old,”

respectively. However, if they determined that the item had been presented but could not remember which voice it had been presented in, they were to press a button simply labeled "old." The labels on the response boxes were changed for each group of subjects.

Procedure

Subjects were tested in groups of six or fewer in a quiet testing room. Stimuli were presented over matched and calibrated TDH-39 headphones at approximately 75 dB SPL. A PDP-11/34 computer was used to present stimuli. The digitized stimuli were converted to 10-kHz 12-bit resolution files for presentation.

Subjects were told that they would be participating in a memory experiment in which they would hear lists of spoken words over the headphones. They were told that after each list was presented they would hear a tone or knock (with examples given), that would indicate whether they should recall the list or perform some math problems. Subjects were instructed to listen carefully to each list because the signal for the task would not occur until after each list was presented; therefore, subjects were unaware until the end of the list whether they would be required to recall the items. The interstimulus interval was 1.5 seconds within lists. Subjects were given 1 minute after presentation of the signal to either recall items or do math problems. Each of these tasks was performed on a piece of paper supplied by the experimenter. After 1 minute, a tone occurred and subjects were instructed to turn over their response sheets so they were no longer in view and prepare for the next word list.

The recognition test occurred about five minutes after the last test or math period. During this time, subjects were given instructions about making old and new judgments. They were told that they would see one item at a time presented on a CRT screen and that they would be required to make judgments on each item. If the item was not presented in the previous study phase, subjects were to press the "new" button. If the item was presented in the previous study phase and the subject could not remember which voice it was presented in, they were told to press the "old" button; however, if they remembered the voice the item was presented in they were told to press the "male-old" or "female-old" buttons.

The recognition test was composed of 96 items, 48 of which had been studied and 48 of which had not. The 48 studied items were obtained by selecting three items from each of the 16 presented lists (always in the serial position 1, 8, and 10). The nonstudied items consisted of the 24 critical lures from all 24 lists (16 studied, 8 not studied) and the 24 items from the nonstudied list (again always in the serial position 1, 8, and 10). The 96 items were randomly presented for each group of subjects. The lists were counterbalanced by having each set of lists serve in each of the three conditions (study-recall, study-math, and nonstudied) across subjects.

Results

Recall

Subjects recalled the critical nonpresented item on 39% of the lists. The serial position curve for studied words is shown in Figure 1. Subjects recalled the critical nonpresented items at about the same rate of studied items presented in the middle of the lists. The false recall in this Experiment was not as high as in Roediger and McDermott's (1995) Experiment 2, perhaps because the test period was only 1 minute in this experiment but was 1.5 minutes in Roediger and McDermott. Still, false recall was robust and approximated recall of words from the middle of the list.

 Insert Figure 1 about here

Recognition

The recognition results are presented in Table 1. The hit rate for the study-recall condition (77%) was not significantly different than the rate observed in the study-arithmetic condition (70%), $t(31) = 1.942$, $p = .06$. The false alarm rate for items from the nonstudied list was .19.

 Insert Table 1 about here

The recognition results for the critical nonpresented lures are shown at the bottom of Table 1. The recognition for the critical lures is even greater than recognition of presented items, however, this was not statistically significant. Once again there is no difference between the hit rates between recalled lists and arithmetic lists, $t(31) = .955$, $p > .1$.

False recognition was actually somewhat higher than correct recognition; that is, the false alarm rate for critical items was 6% higher than the hit rate, $t(31) = 2.09$, $SEM = .028$, $p < .05$. Thus, although false recall was a bit lower than the levels reported by Roediger and McDermott (1995), the level of false recognition was remarkably similar overall.

The primary new question asked in this experiment is whether subjects would be as willing to assign a voice of the speaker to items that were never presented as to items that were presented. In general, the rates of voice attribution are rather similar, although they are a bit higher for studied items than for critical lures. In the Study + Recall condition, subjects provided a voice attribution on 51% of the "old" judgments for studied items, whereas the figure for critical lures was 42%, $t(31) = 1.806$, $SEM = .067$, $p = .08$. The comparable figures for the Study + Arithmetic condition were .44 and .37, $t(31) = 1.983$, $SEM = .071$, $p = .06$, for the studied items and the critical lures, respectively.

A nonsignificant trend was observed for studied items to have slightly greater rates of voice attribution than nonstudied items; however, subjects would quite frequently provide voice attributions to the critical nonstudied items. This evidence fits with other data described in the introduction showing that subjects' false recollections in this paradigm seem quite similar to their veridical recollections based on presented items.

An analysis on the overall effect of prior recall on recognition reveals a positive effect of recall (79% vs. 74%), $t(31) = 2.290$, $SEM = .026$, $p < .05$, on later recognition. In other words subjects have higher recognition rates when subjected to a recall task. However the effect only approached significance for the study + recall (77%) vs. the study + math (70%) conditions $t(31) = 1.942$, $SEM = .038$, $p = .06$; and was not significant for the critical lure + recall (81%) vs. critical lure + math (78%) conditions $t(31) = .955$, $SEM = .035$, n.s. The overall effect is consistent with previous studies on illusory memory reported by Roediger and McDermott (1995).

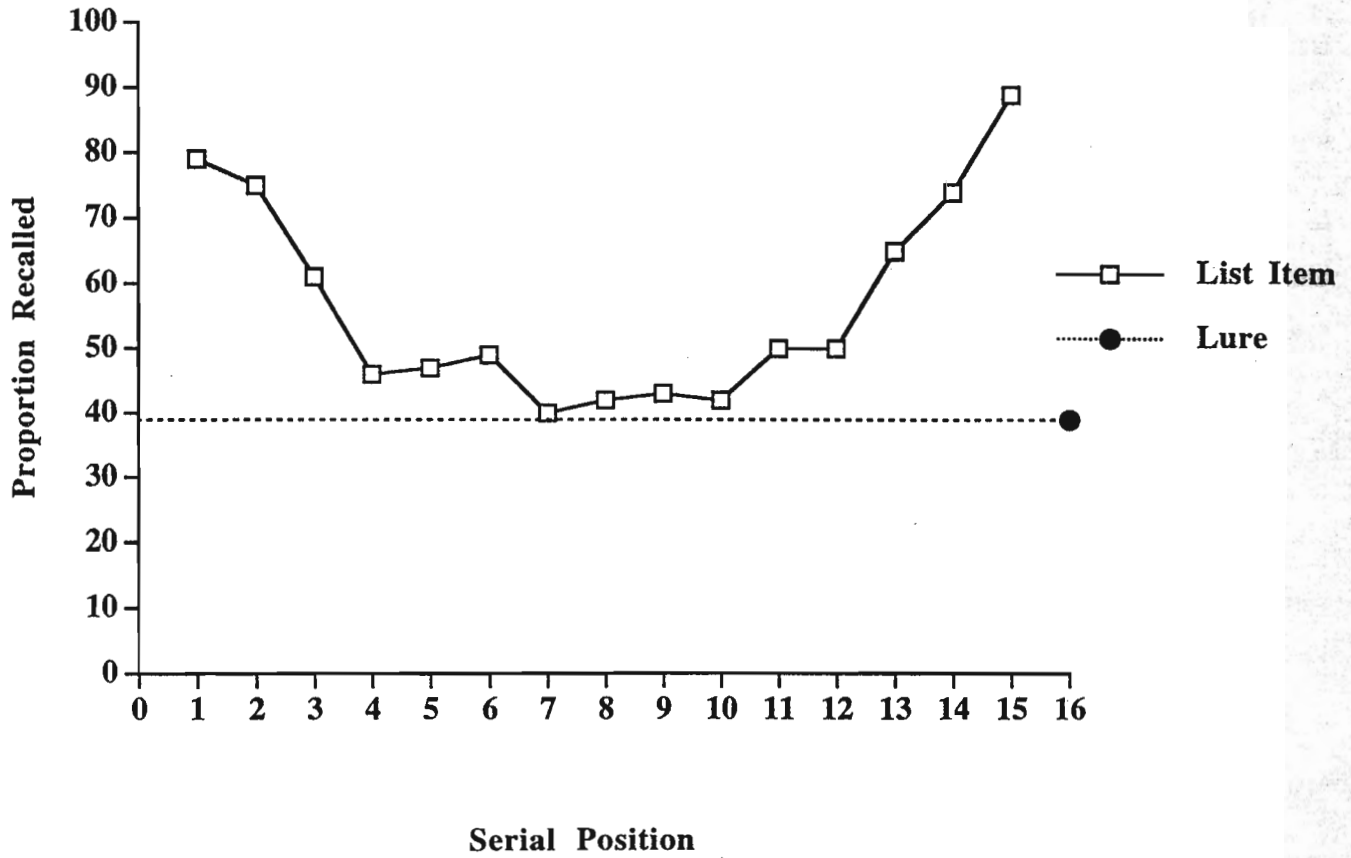


Figure 1. Serial position curve for recall the condition in Experiment 1.

Table 1

Recognition results for studied items (top) and critical lures (bottom) for Experiment 1.
M and F represent male and female judgments respectively.

Item Type	Condition	Proportion of Old Responses				
		Overall	Male	Female	Don't Know	%Voice Attribution
Studied						
	Study +Recall	.77	.21	.18	.38	.51
	Study + Math	.70	.17	.14	.39	.44
	Non-Studied	.19	.04	.03	.12	.37
Critical Lure						
	Study +Recall	.81	.16	.18	.47	.42
	Study + Math	.78	.14	.15	.49	.37
	Non-Studied	.19	.02	.03	.14	.26

Experiment 2

The purpose of the second experiment was to replicate and extend our new findings about recognition of source attributes. The same general procedures were used in Experiment 2 with one exception: subjects were presented with an auditory rather than visual recognition test. Further, the voice in which the test item was presented could either match or mismatch the voice used in the study phase. We were interested in seeing if supplying subjects with voice information during the test might reduce their overall false alarm rates to the critical nonstudied items and also reduce the false attributions of voice to the critical nonstudied words.

Method

Subjects

Forty-four Indiana University undergraduates participated in a one-hour session in partial fulfillment of an introductory psychology course requirement. All of the subjects were native English speakers, and reported no history of speech or hearing impairments at the time of testing. The data from four of the subjects were omitted from the final analysis because they failed to follow instructions.

Materials

The speech materials utilized in the present experiment were identical to the first experiment.

Design

The study and recall phase of Experiment 2 was identical to that of Experiment 1; however, the recognition phase differed in that subjects were presented with auditory test items. The recognition test was composed of 96 items, 48 of which had been studied and 48 of which had not. The 48 studied items were obtained by selecting three items from each of the 16 presented lists (always in the serial position 1, 7, and 10). The nonstudied items consisted of the 24 critical lures from all 24 lists (16 studied, 8 not studied) and the 24 items from the nonstudied list (again always in the serial position 1, 7, and 10). The test items were counterbalanced so that half of the items were presented in a male voice and half in a female voice, and, of those, half were in the same voice as the test item and half were in a different voice as the test item. Half of the 16 lures were presented in a male voice and half presented in a female voice.

The 96 items were randomly presented to each group of subjects. The lists were counterbalanced by having each set of lists serve in each of the three conditions (study-recall, study-math, and nonstudied) across subjects.

Procedure

The general procedure was the same as Experiment 1, except in this experiment, subjects listened to the recognition test over the headphones. As in Experiment 1, subjects were instructed to determine whether an item had been presented during the initial study phase. If the subject determined that an item was not presented during the initial study phase, they were required to press a button labeled "new." However, if they judged that the item was presented during the initial study phase they had three possible responses. If the subject remembered that the item was presented during the initial study phase but did not remember which voice the item was presented in then they were to push a button labeled "old." If the subject

remembered that the study item was presented in the same voice as the test item then they were to push the button labeled "old-same." If the test item was in a different voice than the study item then they were to push a button labeled "old-different."

Results

Recall

Subjects recalled the critical nonpresented item on 38% of the lists. Subjects recalled the critical nonpresented items at about the same rate as studied items presented in the middle of the lists.

Recognition

The recognition results are presented in Table 2. The overall hit rate for the study-recall condition (75%) was significantly different from the recall in the study-arithmetic condition (70%), $t(39) = 2.29$, $p < .05$. The false alarm rate for the nonstudied items was 15%.

 Insert Table 2 about here

The recognition results for the critical nonpresented lures are shown at the bottom of Table 2. Recognition for the critical lures is even greater than recognition of presented items (78% vs. 73%), and this difference approaches significance, $t(39) = 1.781$ $p = .08$. The difference between recognition of lures between recall and arithmetic was not significant, $t(39) = 1.3$ $p > .1$.

One question of interest in this experiment was whether subjects would be more likely to make a voice attribution when auditory information was available during the recognition phase. We observed a significant difference in the rates of voice attribution between the study + recall items (48%) and the critical lure + recall items (38%), $t(39) = 3.098$, $SEM = .062$, $p < .05$. But no difference was observed between the study + math items (46%) and the critical lure + math items (44%) $t(39) = .324$, $SEM = .065$, *n.s.*

An analysis of the overall effect of prior recall on recognition reveals a positive effect of recall (77% vs. 72%), $t(39) = 2.459$, $SEM = 2.0$, $p < .05$, on later recognition. This replicates the effect from Experiment 1. The effect was only significant for the studied items; study + recall (76%) vs. the study + math (70%) conditions, $t(39) = 2.288$, $SEM = 2.15$, $p < .05$; and not for the critical lure + recall (79%) vs. critical lure + math (76%) conditions, $t(39) = 1.300$, $SEM = 2.8$, *n.s.*

The results from Experiment 2 are not as clear-cut as those obtained in Experiment 1. In general the same pattern of source judgments is observed for the studied items and the critical lures; however, when lists have a recall task associated with them it appears that subjects are more likely to make source attributions to studied items than to critical lures. Another pattern that emerges from this experiment is that when subjects do make a source attribution they tend to have a bias for judging an item as occurring in the same voice as the study item regardless of whether it occurred or not. This might suggest that the voice attributions are the result of some general feeling a familiarity with an item and not the result of an actual recollection of an event.

Table 2

Recognition results for studied items (top) and critical lures (bottom) for Experiment 2.

Item Type	Condition	Proportion of Old Responses				
		Overall	Same	Different	Don't Know	%Voice Attribution
Studied						
	Study + Recall Same	.74	.27	.09	.38	.49
	Study + Recall Diff	.77	.23	.14	.40	.47
	Mean	.76	.25	.12	.39	.48
	Study + Math Same	.72	.24	.08	.40	.45
	Study + Math Diff	.69	.24	.08	.37	.46
	Mean	.70	.24	.08	.38	.46
	Nonstudied	.15	.03	.02	.10	.33
Critical Lure						
	Study + Recall	.79	.21	.09	.49	.38
	Study + Math	.76	.23	.10	.43	.44
	Mean	.78	.22	.10	.46	.41

Experiment 3

The purpose of Experiments 3 and 4 was to ask more specific questions about the source judgments and the recollections of our subjects. In the first two experiments, all of the lists that were presented consisted of mixed voices. This might explain why we did not observe any consistent patterns of source judgments. In the following two experiments, we replicate the previous studies except that now the voices in the lists are blocked (either all male or all female voices). This allows us to look at whether listeners are encoding specific attributes of the list items and whether this affects the recollection of the lures. In general, we would expect that subjects will attribute a source to the lure that is the same as the associated list. In Experiment 3, we presented subjects with a visual recognition test and asked them whether they remembered the source of recognized items. In Experiment 4, we presented subjects with an auditory recognition test and asked them to tell us whether the items were presented in the same or different voices.

Method

Subjects

Thirty-four Indiana University undergraduates participated in a one-hour session in partial fulfillment of an introductory psychology course requirement. All the subjects were native English speakers and reported no speech or hearing impairments at the time of testing. The data from one of the subjects were omitted from the final analysis because of his failure to follow instructions.

Materials

The speech materials utilized in the present experiment were identical to those used in Experiments 1 and 2.

Design

The design of Experiment 3 was identical to that of Experiment 1 with one exception; each of the 16 word lists was presented in either a male voice or a female voice.

Procedure

The instructions were identical to those of Experiment 1. The recognition test occurred about five minutes after the last test or math period. During this time, subjects were given instructions about making old and new judgments. They were told that they would see one item at a time presented on a CRT screen and that they would be required to make judgments on each item. If the item was not presented in the previous study phase, subjects were to press the "new" button. If the item was presented in the previous study phase and the subject could not remember which voice it was presented in they were told to press the "old" button; however, if subjects remembered the voice the item was presented in they were told to press the "male-old" or "female-old" buttons.

The recognition test was composed of 96 items, 48 of which had been studied and 48 of which had not. The 48 studied items were obtained by selecting three items from each of the 16 presented lists (always in the serial position 1, 7, and 10). The nonstudied items consisted of the 24 critical lures from all 24 lists (16 studied, 8 not studied) and the 24 items from the nonstudied list (again always in the serial position 1, 7,

and 10). The 96 items were randomly presented to each group of subjects. The lists were counterbalanced by having each set of lists serve in each of the three conditions (study-recall, study-math, and nonstudied) across subjects.

Results

Recall

Subjects recalled the critical nonpresented item on 29% of the lists.

Recognition

The recognition results are presented in Table 3. The overall hit rate for the study-recall condition (74%) was not significantly different from the hit-rate in the study-arithmetic condition (70%), $t(32) = .807$, $p > .1$. The false alarm rate for the nonstudied items was 16%.

Insert Table 3 about here

The recognition results for the critical nonpresented lures are shown at the bottom of Table 3. The recognition for the critical lures was higher for the study + math condition, 75%, than the study + recall condition, 67%, however, this difference was not statistically significant, $t(32) = 1.646$, $p > .1$.

There was no effect in Experiment 3 of prior recall on recognition $t(32) = .613$, $SEM = 3.7$, n.s. This is inconsistent with the effects found in both Experiment 1 and 2. Those experiments showed a small but significant positive effect of recall on later recognition. It should be noted however, that overall subjects in this experiment did much poorer on the recall portion of the experiment than subjects in the first two experiments. The overall recall rate in Experiment 1 (56%) and Experiment 2 (49%), is much higher than that found in Experiment 3 (40%). This is surprising given that the stimuli for all three experiments were similar (with only the voicing within lists changing). However, there is evidence that additional voicing cues provided within lists can aid in serial recall (see Goldinger et al., 1993).

The finding of most interest in this experiment was that subjects are quite accurate at correctly identifying the source attributes for studied items. In other words, subjects are more likely to attribute a female voice to an item if it was presented in a female voice and a male voice if it was presented in a male voice. This indicates that detailed voice information is being encoded with the specific word in memory. Furthermore, the subjects' voice attributions to critical lures are influenced by the voice of the associated word list. In fact, the results indicate that subjects tend to attribute the voice of the associated list to the critical lure with the same frequency as they attribute the correct source to an actual presented item.

Experiment 4

The results of Experiment 3 demonstrate the subjects can remember specific details about the physical attributes of a word list and that illusory memories are influenced by those specific details. The purpose of Experiment 4 was to investigate the effect that voices have during the recognition phase of the experiment. The general design of Experiment 4 was the same as Experiment 3 with one exception: in Experiment 4, the recognition list was auditory instead of visual. This allows us to look at recognition rates for consistent

Table 3

Recognition results for studied items (top) and critical lures (bottom) for Experiment 3.

Item Type	Condition	Proportion of Old Responses				
		Overall	Male	Female	Don't Know	% Voice Attribution
Studied						
	Study + Recall Male	.73	.22	.09	.42	.42
	Study + Recall Female	.75	.10	.22	.43	.42
	Mean	.74	.16	.16	.43	.42
	Study + Math Male	.70	.18	.07	.45	.36
	Study + Math Female	.70	.06	.19	.45	.36
	Mean	.70	.12	.26	.45	.36
	Nonstudied	.16	.02	.02	.12	.25
Critical Lure						
	Study + Recall Male	.67	.20	.07	.40	.40
	Study + Recall Female	.67	.09	.18	.40	.39
	Mean	.67	.15	.12	.40	.40
	Study + Math Male	.73	.20	.08	.45	.38
	Study + Math Female	.78	.05	.21	.52	.33
	Mean	.75	.12	.14	.49	.35

voice lures versus inconsistent voice lures. Previous research on recognition memory for spoken words has shown that voice information aids in veridical recognition of list items (Palmeri, Goldinger, & Pisoni, 1993). Therefore, we would expect that listeners should be more likely to identify a lure as old if it were presented in the same voice as the associated list.

Method

Subjects

Thirty-two Indiana University undergraduates participated in a one-hour session in partial fulfillment of an introductory psychology course requirement. All of the subjects were native English speakers and reported no speech or hearing impairments at the time of testing. The data from two of the subjects were omitted from the final analysis because they failed to follow instructions.

Materials

The study materials utilized in the present experiment were identical to the previous three experiments. The voices for each of the lists were presented in a blocked format.

Design

The study, recall, and recognition phases of Experiment 4 were identical to those of Experiment 2. The test items were counterbalanced so that half of the items were presented in a male voice and half in a female voice, and, of those, half were in the same voice as the test item and half were in a different voice as the test item. The 16 critical lures were presented in both male and female voices, half, were presented in the same voice as the associated study list and half were presented in a different voice as the associated study list.

Procedure

The procedure for Experiment 4 was identical to that of Experiment 2.

Results

Recall

Subjects recalled the critical nonpresented item 37% of the time. Subjects recalled the critical nonpresented items at about the rate of studied items presented in the middle of the lists.

Recognition

The recognition results are presented in Table 4. The overall hit rate for the study-recall condition (75%) was significantly different from the hit rate observed in the study-arithmetic condition (68%), $t(29) = 2.833$, $p < .05$. The false alarm rate for the nonstudied items was 11%.

Insert Table 4 about here

Table 4

Recognition results for studied items (top) and critical lures (bottom) for Experiment 4.

Item Type	Condition	Proportion of Old Responses				
		Overall	Same	Different	Don't Know	%Voice Attribution
Studied						
	Study + Recall Same	.76	.32	.10	.34	.55
	Study + Recall Diff	.73	.12	.28	.33	.55
	Mean	.75	.22	.19	.34	.55
	Study + Math Same	.69	.30	.07	.32	.54
	Study + Math Diff	.67	.14	.15	.38	.43
	Mean	.68	.22	.11	.35	.49
	Nonstudied	.11	.02	.02	.07	.36
Critical Lure						
	Study + Recall Same	.75	.17	.22	.36	.49
	Study + Recall Diff	.80	.05	.27	.48	.40
	Mean	.78	.11	.25	.42	.46
	Study + Math Same	.74	.21	.17	.36	.51
	Study + Math Diff	.72	.09	.25	.38	.47
	Mean	.73	.15	.21	.37	.49

An analysis on the overall effect of prior recall on recognition reveals a positive effect of recall (76% vs. 70%) $t(29) = 3.118$, $SEM = 2.3$, $p < .01$, on later recognition. The effect was significant only for the studied items; study + recall (75%) vs. the study + math (68%) conditions $t(29) = 2.838$, $SEM = 2.3$, $p < .01$; and not for the critical lure + recall (78%) vs. critical lure + math (73%) conditions $t(29) = 1.408$, $SEM = 3.4$, n.s. This replicates the effect from Experiments 1 and 2, but not Experiment 3. A look at the recall results in this experiment reveals that subjects recall levels (46%) are more similar to the levels observed in Experiments 1 and 2.

The recognition results for the critical nonpresented lures are shown at the bottom of Table 4. The recognition for the critical lures is even greater than the recognition of presented items (75% vs. 71%), however, this difference was not statistically significant, $t(29) = 1.546$, $p > .1$. Of interest, however, is the pattern of responses for the items that were presented in the same voice at study and test. Subjects tended to judge an item as old-same more often than old-different when the studied item and the test item were in the same voice for both the recalled lists (32% vs. 10%), $t(29) = 4.115$, $SEM = 3.5$, $p < .001$ and non-recalled lists (30% vs. 7%), $t(29) = 6.541$, $SEM = 2.75$, $p < .001$. However, this pattern was not observed for the critical lures in either recalled (17% vs. 22%), $t(29) = .865$, $SEM = 3.85$, n.s, or non-recalled lists (21% vs. 17%), $t(29) = .166$, $SEM = 2.5$, n.s. This result suggests that although illusory memories act like veridical memories in many ways, there are important fundamental differences that are assessed with different tasks. It might be the case that illusory memories are based on abstract representations, whereas veridical memories are based on episodic traces of events (Hintzman, 1986, 1988). Further research is necessary in order to determine how these two types of memories differ.

General Discussion

The four experiments reported here provide consistent evidence for false recall and false recognition. A small, but significant, positive effect of previous recall on recognition was observed in three of the four experiments. We also demonstrated that subjects are willing to attribute source judgments to nonpresented items. We discuss these results in relation to previous findings in illusory memory and implication for future research.

First, we obtained false recall effects that ranged from .29 to .38. Although these effects are impressive, the magnitude is slightly below the levels originally reported by Roediger and McDermott (1995) and others more recently (McDermott, 1996; Payne et al., 1996; Norman & Schacter, 1996; Schacter et al., 1996). One possible reason for the lower levels of false recall is that we reduced the amount of recall time after the 8 tested lists in our experiments from 90 seconds to 60 seconds. Because false recall often occurs towards the end of the subjects' responses (Roediger & McDermott, 1995), the shorter recall period may have had the unfortunate effect of reducing both veridical and false recall, but also possibly disproportionately harming the latter quantity. In retrospect, this reduction of recall time was unfortunate, but nonetheless, we did observe relatively high levels of false recall across the four experiments, with the weighted average being 36%. False recall in the Roediger-McDermott (1995) paradigm is robust, especially, in light of instructions to subjects not to guess and of the generally low level of other intrusions obtained in these experiments.

A second issue examined in the present experiments was whether recall (of studied and nonstudied items) affected later recognition. In general, the answer is yes, confirming the results of Roediger and McDermott (1995). This was the case in three of the four experiments. It should also be noted that overall recall in Experiment 3 (where no effect was observed) was greatly reduced compared to recall in the other three experiments. This low level of recall might account for the lack of an effect on recognition. A meta

analysis over all four experiments on the factors of recall (recalled x not recalled) and study (studied x not studied), revealed only a marginal effect of recall, $F(1,536) = 3.31$, $p = .06$. In summary, prior recall of a list seems to have a moderate but relatively consistent effect on later recognition. In addition, McDermott (1996) has shown that three tests have a more powerful effect than one test in boosting later recall of both studied and nonstudied items. The act of retrieval does seem to play a critical role in the development of false recall and probably of false recognition, too (see Roediger, et al., in press, for further elaboration).

The present experiments were specifically designed to give us further insight into our subjects' phenomenological experience during false recognition. Overall we found that listeners were as likely to attribute a voice to an illusory memory as a veridical memory. Furthermore, an analysis over all four experiments, reveals that subjects are more likely to label as "old" a critical lure than an actual studied item, $F(1,536) = 4.8$, $p < .05$. In general, false recognition was equal to or exceeded veridical recognition and the source judgments occurred with about the same frequency for both types of items. Further, this pattern held up over manipulations at encoding (presenting lists purely in one voice or mixed, within-list, as to voice) and over manipulations at retrieval (testing either with visual or auditory recognition tests). The finding that changing neither encoding nor retrieval factors greatly affected the levels of false recognition or voice attribution shows how powerful the memory illusion arising in the Roediger-McDermott (1995) paradigm is at promoting false recognitions. One might expect that subjects would be able to use differences in voice information (either when whole lists are presented in one or the other voice, or when test items are presented in the same or a different voice rather than being visually presented) as powerful retrieval cues to aid recognition. Yet, subjects apparently cannot use this information to lower levels of false recognition. However, it should be noted here that although voice information did not change the overall level of illusory recognition, patterns emerged which indicated that there is something fundamentally different between illusory and veridical memories. This was most evident in comparisons of Experiments 3 and 4. The results from Experiment 3 demonstrate that subjects' illusory memories are influenced by physical attributes of the associated list. Subjects in the third experiment tended to identify the voice of the nonpresented item to be the same as the associated word list. However, the results from Experiment 4 show that when faced with an auditory recognition test subjects were less likely to make the same attribution. Further research is necessary to tease apart the differences between false and veridical recognition.

The levels of voice attribution observed here (voices were attributed for 30-40% of items called old) were much lower than those reported by Payne et al. (1996). However, as noted previously, Payne et al.'s paradigm differed considerably from the present recognition procedure, because they had subjects make judgments after the third of three successive recall tests and to respond to each recalled item by responding male, female or don't know to indicate their source attributions. In addition, our experimental procedures may have biased subjects to place their response criterion more conservatively to saying that an item was heard in a particular voice. We gave subjects the option of responding "old" with no voice attribution and it may have been that subjects used this option unless they were highly confident that they remembered the voice of presentation. Other testing procedures, such as having subjects always respond "old" for recognized items and then always to give the judgment of male, female, or don't know might cause them to be more willing to provide judgments of voice. This procedure would more closely approximate the methods used by Payne et al. We plan to adopt these procedures in future recognition experiments dealing with source attributions of voice of speaker.

Taken together with prior results, the present findings replicate and extend the conclusions that subjects' false recall and false recognition in the Roediger-McDermott (1995) paradigm are robust and that these false responses have, to the subjects, the same phenomenological characteristics as veridical recall and recognition of studied material. False recollections are often judged as remembered (Payne et al., 1996;

Roediger & McDermott, 1995) when Tulving's remember/know paradigm is used, meaning that subjects believe they remember features about the actual occurrence of test items that were not presented. Further, Norman and Schacter (1996) have shown recently that even when subjects are asked to write down some feature they remember from the study event, the levels of false recognition do not drop (relative to a condition in which subjects did not have to justify remember judgments in this way). In our experiments, and those of Payne et al. (1996), subjects attribute a presentation voice to critical lures at the same general level as actual presented items. This suggests that, at the very least, subjects retained a general knowledge about base rates and frequency of occurrence of different voices during study. The challenge for future studies remains to specify why these items that are not presented are remembered as well as items that are presented.

References

- Bartlett, F.C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: The University Press.
- Goldinger, S.D. (1992). Words and voices: Implicit and explicit memory for spoken words. *Research in Speech Perception Technical Report No. 7*, Indiana University, Bloomington, IN.
- Hintzman, D.L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.
- Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.
- Jones, T.C. & Roediger, H.L. (1995). The experiential basis of serial position effects. *European Journal of Cognitive Psychology*, **7**, 65-80.
- Lockhart, R. (1975). The facilitation of recognition by recall. *Journal of Verbal Learning and Verbal Behavior*, **14**, 253-258.
- McDermott, K.B. (1996) The persistence of false memories in list recall. *Journal of Memory and Language*, **35**, 212-230.
- Norman, K. & Schacter, D.L. (1996). False recognition in younger and older adults: Exploring the characteristics of illusory memories. Manuscript submitted for publication.
- Palmeri, T.J., Goldinger, S.D. & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **19**, 309-328.
- 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 & Review*, **3**, 105-111.
- Roediger, H.L. & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.
- Roediger, H.L., McDermott, K.B., & Goff, L. (in press). Recovery of true and false memories: Paradoxical effects of repeated testing. In M.A. Conway (Ed.) *Recovered memories and false memories*. Oxford: Oxford University Press.
- Russell, W.A., & Jenkins, J.J. (1954). *The Complete Minnesota Norms for Responses to 100 Words from the Kent-Rosanoff Word Association Test* (Technical Report No. 11, Contract N8 ONR 66216). University of Minnesota.

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.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, **62**, 1-12.