



Journal of Articles in Support of the Null Hypothesis

Vol. 3, No. 1

Copyright 2004 by Reysen Group. 1539-8714

www.jasnh.com

False Recall Does Not Increase When Words are Presented in a Gender-Congruent Voice

David S. Kreiner

R. Zane Price

Amy M. Gross

Kristy L. Appleby

Central Missouri State University

We investigated how false recall of words might be affected by the consistency of the gender of the person speaking the words with the gender of the person listening to the words. Ninety-eight college students listened to eight 15-item word lists presented in a male or female voice. A non-presented lure word was associated with all the presented words on each list. Participants recalled a mean of 3.30 (SD = 1.83) non-presented lure words out of eight possible, for a false recall rate of 41%. False recall was not significantly affected by either speaker gender or listener gender. Contrary to the hypothesis, no interaction was observed between the gender of the speaker and the gender of the listener.

It is now well established that, under certain conditions, people recall hearing words which were not in fact presented to them. Roediger and McDermott (1995) presented word lists to college students and then asked them to recall the words they had heard. Each list contained 12 words, all associated with a target word that was not presented (henceforth referred to as the critical lure). For example, words associated with “chair,” such as “table,” “sit,” and “legs” were presented, but the critical lure, “chair,” was not. After listening to each of six such lists, the students were asked to write down all of the words they recalled. The critical lure (e.g., “chair”) was recalled 40% of the time despite the fact that it had not been presented. In a recognition test, the students recognized the critical lures as “old” items 84% of the time. In a second experiment, using 15-word lists, Roediger and McDermott reported a false recall rate of 55% for the critical lures and a false recognition rate of 81%. This method is known as the Deese-Roediger-McDermott (DRM) paradigm, and the findings have been replicated numerous times (Roediger, Watson, McDermott, & Gallo, 2001).

The probability of falsely recalling critical lures varies considerably. Much of the list-to-list variability in false recall can be explained by list recallability and associative connections from list words to the critical lure (Roediger et al., 2001). However, there has been little investigation of how false recall in this paradigm is related to individual difference factors, such as gender.

Women tend to outperform men on measures of episodic memory ability (Herlitz, Airaksinen, & Nordstrom, 1999; Lewin, Wolgers, & Herlitz, 2001), and because the Deese-Roediger-McDermott paradigm measures episodic memory for the presentation of particular words, we might expect women to be more accurate at recalling which words were presented.

Seamon, Guerry, Marsh, and Tracy (2002) found no gender difference in false recall using the DRM paradigm. This failure to find a gender difference may be due to the fact that the DRM paradigm relies on episodic memory for words, and although there is evidence that women outperform men on tests of verbal memory (Kimura & Clarke, 2002), gender differences in verbal ability appears to be small (Hyde & Linn, 1988). Herlitz et al. (1999) reported that the gender difference in episodic memory could not be explained completely by differences in verbal ability.

Gender may play a role in false recall even if there is no overall gender difference. The purpose of the present study was to determine if false recall of words could be increased by similarity between the voice of the person presenting the words and the internal voice of the listener. In particular, we investigate whether false recall is more likely when the word lists are presented in a voice matching the gender of the listener than when the words are presented in an opposite gender voice. Hearing words presented in a gender-matching voice may make it difficult to discriminate between items that the listener actually heard and items that the listener merely thought.

Non-presented words may be recalled falsely when the individual thinks of the word but fails to recall that it was not presented by the experimenter (Bredart, 2000). When individuals do correctly discriminate between words that were presented and words that are similar but were not presented, it may be the result of a process that suppresses a signal indicating high similarity between the associated word and the presented items. A similarity explanation of false recall was also provided by Underwood (1965), who provided evidence of an “implicit associative response,” meaning that a stimulus word such as “butter” generates a particular associated word such as “bread,” which the listener later falsely recalls as having been presented.

Schacter et al. (1996) suggest that the right frontal lobe of the brain is important suppressing similarity to avoid false recall. Curran, Schacter, Norman, and Galluccio (1997) reported a case of a man with a lesion to the right frontal lobe who showed abnormally high levels of false recall and appeared to rely heavily on similarity. This case is consistent with Schacter et al.'s (1996) suggestion that a right frontal lobe process serves to suppress an over-reliance on similarity in making judgments about which items were presented and which were not.

False recall should therefore be more likely in situations in which there is high similarity between presented and non-presented items and in cases in which this similarity is difficult to suppress. It may be particularly difficult to determine which words entered one's thoughts but were not actually presented if the voice in which the words are presented is similar to one's own internal voice. For example, a man who listens to a list of associated words that are all associated with chair may think of the word "chair" even though it was not presented. When the man later attempts to recall the words he heard, he must determine whether he actually heard "chair" or whether he only thought of it. This may be more difficult if he heard the word list presented in a man's voice than if the word list had been presented in a woman's voice.

This explanation is consistent with the source-monitoring framework. Source monitoring is defined as "the set of processes involved in making attributions about the origins of memories, knowledge, and beliefs" (Johnson, Hashtroudi, & Lindsay, 1993, p. 3). An important aspect of source monitoring is attempting to determine whether a particular memory is the result of simply thinking about an event, or whether the memory is the result of an external event (i.e., the presentation of a particular word). Deficits in source monitoring can result from frontal lobe damage (Johnson, Hashtroudi, & Lindsay, 1993), consistent with the evidence from Schacter et

al. (1996) that frontal lobe processes are important in suppressing similarity when deciding whether an item was previously presented or not.

The source-monitoring framework holds that source memory can vary on a continuum from less precise to more precise. Dodson, Holland, and Shimamura (1998) demonstrated this aspect of source monitoring by presenting words spoken by two males and two females and then asking listeners to recall the gender of the person who spoke the words in addition to identifying the specific speaker. They found that listeners could often determine whether the speaker was male or female even when they could not identify the specific speaker. This is an example of partial source memory, as the listeners recalled some information about the presentation of the words but could not recall the specific source. Dodson et al. (1998) found that a divided attention task (remembering five digits) made it more difficult to identify the specific speaker but did not make it more difficult to identify the gender of the speaker. This suggests that speaker gender is a robust form of partial source memory. In similar vein, Palmeri, Goldinger, and Pisoni (1993) found that, when asked to judge whether words had been presented earlier in the same voice or in a different voice, listeners tended to perceive words presented in a same-gender voice as having been presented in the same voice, even though the word had in fact been presented in two different voices (i.e., two different female voices or two different male voices). This supports the view that gender of the speaker is an important aspect in how people perceive similarity of voices.

Parks (1997) provides several demonstrations of false recall in which individuals remember saying things that they did not say. In one study, participants were shown a series of cards containing phrases, some of which they were instructed to say aloud. They were then given a recognition test that required them to choose which of two similar phrases had

presented. The participants remembered saying 16% of the phrases that they had in fact not said aloud. In another study, participants were presented with cards containing two questions, one of which they were instructed to answer aloud. Parks (1997) reported that the participants falsely recalled answering 22% of the questions that they had not answered aloud. These demonstrations may be taken as evidence for difficulty in monitoring the source of the phrases; the students evidently remembered saying things that they had merely thought about saying. This suggests that one's internal voice may be confused with one's external voice in source monitoring.

We hypothesized that this difficulty in identifying words that were not presented would be related to the gender match or mismatch between the speaker and the listener. The probability of falsely recalling an associated but non-presented word is expected to be greater when there is a match between the gender of the speaker and the gender of the listener: men should show higher false recall after listening to a word list presented in a man's voice and women should show higher false recall after listening to a word list presented in a woman's voice. This would appear as a significant interaction between the speaker gender and listener gender in a factorial design.

Method

Participants

Ninety-eight students at Central Missouri State University enrolled in Psychology and Geography courses volunteered to participate. Many of the students received extra credit or met one of several course requirements by participating. The sample included 49 women and 49 men. (Recruitment of participants continued until equal numbers of each gender had participated). Age of the participants ranged from 18 to 46 years ($M = 22.54$, $SD = 5.24$).

Materials & Designs

We used a 2 x 2 between subjects factorial design, with the independent variables being participant gender and speaker gender. A man and a woman from the university community, both who had clear speaking voices and were unaware of the hypothesis, volunteered to record the word lists. We selected eight word lists from Roediger and McDermott (1995), specifically the lists for the following critical lures: bread, doctor, girl, man, needle, sleep, spider, and window. The lists were recorded with a one-second pause between each word and a 2.5 minute pause for recall between word lists. The word “recall” was recorded at the end of each word list. A portable compact disc player was used to present the recorded word lists. Participants recorded their responses on response sheets that contained 16 blanks for each word list. Each set of blanks was headed with the number of the word list (List 1, List 2, etc.), and blanks for two words lists were printed on each page. We used the Edinburgh Handedness Inventory (Oldfield, 1971) to obtain a quantitative measure of handedness.

Procedure

After an informed consent procedure, we gave participants the following instructions:

Please listen carefully to each list of words. Do not write anything while listening to the word lists. At the end of each list, you will hear the word “Recall.” At that point, try to remember and write down as many of the words as you can. You should start by trying to remember the last few words on the list, but it is okay to write the words in any order. Do not worry about spelling.

Students participated in small groups, with the gender that the words were presented in randomly assigned to testing sessions. Each participant listened to all eight word lists presented in the same voice (either male or female). The word lists were presented in the same order for all

participants. Participants recorded their responses on the response sheets provided for them. They had continual access to the response sheets as the word lists were presented, and were not prevented from looking at the sheets.

We measured the number of (non-presented) critical lure words recalled, the number of presented words correctly recalled, and the number of intrusions recalled (non-presented words other than lure words). Recognizable misspellings of words were counted as having been recalled.

Results

Overall, participants recalled a mean of 3.30 ($SD = 1.83$) lure words out of eight possible, indicating a false recall rate of 41%. The mean number of presented words recalled was 67.32 ($SD = 10.79$) out of a possible 120, and the mean number of intrusions was 2.64 ($SD = 2.54$). Two-way between subjects ANOVAs were conducted for each of the three dependent variables: number of lure words recalled, number of presented words recalled, and number of intrusions. We used the .05 alpha level for all tests of significance, and we report eta-squared (η^2) as a measure of effect size with each F -test. The *a priori power* for each effect, calculated using G-Power (Buchner, Faul, & Erdfelder, 1996) was .69 for each effect, based on a medium effect size ($f = .25$).

Means by speaker and listener gender for the numbers of lure words recalled are presented in Table 1. The effect of speaker gender was not significant, $F(1,94) = 0.77$, $p = .38$, $\eta^2 = .01$. Women tended to recall more lure words than men, but the effect size was small and the difference did not reach statistical significance, $F(1,94) = 3.73$, $p = .06$, $\eta^2 = .04$. There was no interaction of speaker gender with listener gender, $F(1,94) = 0.03$, $p = .85$, $\eta^2 = .00$.

Means for the numbers of presented words recalled are presented in Table 2. The effect of speaker gender was not significant, $F(1,94) = 1.42, p = .24, \eta^2 = .01$. Women recalled a mean of 6.51 more words than men, a statistically significant difference $F(1,94) = 8.69, p = .004, \eta^2 = .08$. Speaker and listener gender did not show an interaction, $F(1,94) = 0.10, p = .75, \eta^2 = .00$.

Means numbers of intrusions are shown in Table 3. There was not a significant effect of speaker gender, $F(1,94) = 0.92, p = .34, \eta^2 = .00$, listener gender, $F(1,94) = 0.15, p = .70, \eta^2 = .01$, or a significant interaction of speaker and listener gender, $F(1,94) < .01, \eta^2 = .00$. Intrusions tended to be words associated with the list. For example, the presented items on the list for the critical lure “bread” included: butter, food, eat, sandwich, rye, jam, milk, flour, jelly, dough, crust, slice, wine, loaf, and toast. Intrusions for that list included: wheat, water, roll, piece, knife, and cheese.

Handedness scores on the Edinburgh Inventory ranged from -89 to 100 ($M = 57.13, SD = 43.46$). Negative scores indicate greater left hand dominance and positive scores indicate greater right hand dominance (Oldfield, 1971). Handedness scores were not significantly correlated with the number of lure words recalled, $r(96) = -.06, p = .53$, the number of presented words recalled $r(96) = .08, p = .46$, or the number of intrusions, $r(96) = -.06, p = .58$.

Discussion

The results did not support the hypothesis that presenting words in a gender-matching voice increases the probability of incorrectly recalling words that were not presented. Speaker and listener gender did not have interactive effects on number of critical lures recalled, number of presented words recalled, or number of intrusions. The effect sizes for the interaction of speaker and listener gender were essentially zero for all three dependent variables. This suggests that the failure to find an interaction in the present study was not simply a case of an effect that

did not reach statistical significance due to a lack of power, but rather that there was no interaction effect at all.

Similarly, there was no significant effect of speaker gender on any of the three dependent variables, with effect sizes close to zero. While no such effect was hypothesized, these results indicate that the two individuals who recorded the word lists did not differ on the intelligibility of their pronunciations. Had there been an effect of speaker gender, it would have been difficult for us to determine whether it was due to a general effect of the gender of the person speaking or to an idiosyncratic difference between our two volunteer speakers.

There was little evidence for a gender difference in false memory. While women recalled a mean of .67 more critical lure words than men (out of eight possible), the difference was not statistically significant and the effect size was very small, explaining only four percent of the variability in false recall. The lack of a gender difference in false recall is consistent with previous findings (Seamon et al., 2000). Although women showed greater recall of presented words than men, the effect size was small. This small gender difference in recall is consistent with previous research showing small gender differences in verbal ability (Hyde & Linn, 1988).

The results of the present study suggest that individuals are not any more likely to falsely recall words when they are presented in a voice matching their own gender. This may indicate simply that gender matching of voices is not a major factor in terms of similarity. It may still be the case that similarity of one's own voice to the voice one hears is a factor in false recall, consistent with the source monitoring framework. The ideal method to test this hypothesis would be to measure false recall when the stimuli are presented in the *listener's* voice as compared to another person's voice. Of course, this would create a serious methodological problem as the listener would have heard all of the stimuli items while speaking them. Another approach would

be to carefully manipulate the voice in which the words are presented in terms of factors such as frequency, prosody, and regional accents. This line of research is important in determining how interactions between individual differences and situational factors may contribute to false recall.

References

- Bredart, S. (2000). When false memories do not occur: Not thinking of the lure or remembering that it was not heard? *Memory, 8*, 123-128.
- Buchner, A., Faul, F., & Erdfelder, E. (1996). G-Power: A priori, post-hoc, and compromise power analyses for the Macintosh (Version 2.1). [Computer Program]. Trier, Germany: University of Trier.
- Curran, T., Schacter, D.L., Norman, K.A., & Galluccio, L. (1997). False recognition after a right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia, 35*, 1035-1049.
- Dodson, C.S., Holland, P.W. , & Shimamura, A.P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1121-1136.
- Herlitz, A., Airaksinen, E., & Nordstrom, E. (1999). Sex differences in episodic memory: The impact of verbal and visuospatial ability. *Neuropsychology, 13*, 590-597.
- Hyde, J.S., & Linn, M.C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53-69.
- Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.
- Kimura, D., & Clarke, P.G. (2002). Women's advantage on verbal memory is not restricted to concrete words. *Psychological Reports, 91*, 1137-1142.
- Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. *Neuropsychology, 15*, 165-173

Send Manuscript Correspondence to

Theresa Kay

Psychology Department

- Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 91-113.
- 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, and Cognition*, *19*, 309-328.
- Parks, T.E. (1997). False memories of having said the unsaid: Some new demonstrations. *Applied Cognitive Psychology*, *11*, 485-494.
- Roediger, H.L., & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *4*, 803-814.
- Roediger, H.L., Watson, J.M., McDermott, K.B., & Gallo, D.A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin and Review*, *8*, 385-407.
- Schacter, D.L., Reiman, E., Curran, T., Yun, L.S., Bandy, D., McDermott, K., & Roediger, H.L. (1996). Neuroanatomical correlates of veridical and illusory recognition memory revealed by PET. *Neuron*, *17*, 267-274.
- Seamon, J.G., Guerry, J.D., Marsh, G.P., & Tracy, M.C. (2002). Accurate and false recall in the Deese/Roediger and McDermott procedure: A methodological note on the sex of the participant. *Psychological Reports*, *91*, 423-427.
- Underwood, B.J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122-129

Author Notes

Correspondence concerning this article may be sent to David S. Kreiner, Department of Psychology, Central Missouri State University, Warrensburg, MO, 64093, email kreiner@cmsu1.cmsu.edu.

We are grateful to Vicki Silvers-Gier and to Byron Johnson at KTBG for recording the word lists.

This research was presented at the 2003 Annual Meeting of the Psychonomic Society in Vancouver B.C.

Table 1*Mean Numbers of Lure Words Recalled by Speaker and Listener Gender*

Speaker Gender	Listener Gender						Marginals	
	Female (n=49)			Male (n=49)			M	SD
	M	SD	n	M	SD	n		
Female (n=50)	3.46	2.10	28	2.82	1.94	22	3.18	1.60
Male (n=48)	3.89	1.82	21	3.07	1.33	27	3.42	1.60
Marginals	3.63	1.98		2.96	1.62			

Table 2*Mean Numbers of Presented Words Recalled by Speaker and Listener Gender*

Speaker Gender	Listener Gender						Marginals	
	Female (n=49)			Male (n=49)			M	SD
	M	SD	n	M	SD	n		
Female (n=50)	71.36	11.78	28	65.82	9.93	22	68.92	11.25
Male (n=48)	69.52	10.26	21	62.63	9.12	27	65.65	10.13
Marginals	70.57	11.08		64.06	9.53			

Table 3*Mean Numbers of Intrusions by Speaker and Listener Gender*

Speaker Gender	Listener Gender						Marginals	
	Female (n=49)			Male (n=49)			M	SD
	M	SD	n	M	SD	n		
Female (n=50)	3.00	2.11	28	2.77	3.18	22	2.90	2.60
Male (n=48)	2.48	2.31	21	2.30	2.63	27	2.38	2.47
Marginals	2.78	2.19		2.51	2.87			

*Received: January 5, 2004**Revised: June 11, 2004**Accepted: July 5, 2004*