Memory & Cognition
1999, 27 (3), 494-500

Norms for word lists that create false memories

MICHAEL A. STADLER
University of Missouri, Columbia, Missouri

and

HENRY L. ROEDIGER III and KATHLEEN B. McDERMOTT
Washington University, St. Louis, Missouri

Roediger and McDermott (1995) induced false recall and false recognition for words that were not presented in lists. They had subjects study 24 lists of 15 words that were associates of a common word (called the critical target or critical lure) that was not presented in the list. False recall and false recognition of the critical target occurred frequently in response to these lists. The purpose of the current work was to provide a set of normative data for the lists Roediger and McDermott used and for 12 others developed more recently. We tested false recall and false recognition for critical targets from 36 lists. Despite the fact that all lists were constructed to produce false remembering, the diversity in their effectiveness was large—60% or more of subjects falsely recalled *window* and *sleep* following the appropriate lists, and false recognition for these items was greater than 80%. However, the list generated from *king* led to 10% false recall and 27% false recognition. Possible reasons for these wide differences in effectiveness of the lists are discussed. These norms serve as a useful benchmark for designing experiments about false recall and false recognition in this paradigm.

Roediger and McDermott (1995) designed experiments to study false recall and false recognition that were based on a technique first used by Deese (1959b). In a series of experiments in the late 1950s, Deese (1959a, 1959b) was interested in learning how associative factors affected recall. Deese (1959a) presented subjects with 15-word lists that varied in their interitem associative strength, "defined as the average relative frequency with which all items in a list tend to elicit all other items in the same list as free associates" (p. 305). He showed that this measure correlated highly (+.88) with the number of items recalled from the list but negatively with the number of extralist intrusions that subjects produced (−.48). However, the stronger the associative bonds between list items, the more likely were subjects to produce the same common associate as an intrusion (+.55). In summary, lists of words that were strongly interassociated tended to produce accurate recall; when an intrusion did occur for these lists, it was likely to be highly similar among subjects.

The authors are very grateful to Tim Capstick, Julianna Gooch, Shana Lesch, Rachel Stanovcic, Justin Warren, and Erica Woodard for completing the monumental task of scoring these data and entering them into electronic format. We also thank David Payne, Don Read, and Gene Winograd for their comments on a previous draft of this paper. Correspondence concerning this article should be addressed to M. A. Stadler, Department of Psychology, University of Missouri, 210 McAlester Hall, Columbia, MO 65202; H. L. Roediger III, Department of Psychology, Box 1125, Washington University, One Brookings Drive, St. Louis, MO 63130-4899; or K. B. McDermott, Division of Radiological Sciences, 4525 Scott Ave., Washington University School of Medicine, Box 8225, St. Louis, MO 63130-1093. Electronic mail may be sent to the authors at stadlerm@.missouri.edu, roediger@artsci.wustl.edu, or kmcd@npg.wustl.edu, respectively.

To examine their tendency to produce intrusions, Deese (1959b) presented 50 subjects with 12-word lists that varied in their interitem associative strength. For example, he presented subjects with the 12 most common associates to the word *butterfly* (*moth*, *insect*, *wing*, *bird*, *fly*, *yellow*, *net*, *pretty*, *flower*, *bug*, *cocoon*, *color*) and asked them to recall the list in any order (single trial, free recall). In this instance, not 1 subject incorrectly recalled *butterfly*. However, for other lists constructed in exactly the same way, intrusions were quite high. For example, when given the 12 most common associates to the word *needle*, 42% of the subjects intruded the critical nonpresented word in immediate free recall.

Deese's (1959b) report examining intrusions to strongly associated lists of words rested quietly in the literature, known to only a few researchers until Roediger and McDermott (1995) revived it. Because many of the lists (such as the butterfly list) did not produce many intrusions, and because the more famous work by Deese (1959a, 1965) emphasized the power of associations in producing high levels of veridical recall, the possibility of using this technique to produce false memories was overlooked. Indeed, because single-trial free recall usually leads to very small numbers of intrusions, especially with unrelated word lists (see Cofer, 1967; Roediger & Payne, 1985), there was every reason to regard Deese's (1959b) findings of high intrusion rates with only a few of his lists with some suspicion. However, the technique is now receiving much more attention. In an interesting historical review, Bruce and Winograd (1998) argue that the different reactions in the scientific community to the Deese (1959b) and Roediger and McDermott (1995) papers offer a classic illustration of the zeitgeist's influence in science. Basically, in 1959, errors

in memory were regarded mostly as a nuisance, and so received little attention; by the early 1990s, the genesis of memory errors was one of the main issues in the memory literature (see Roediger, 1996).

Roediger and McDermott (1995, Experiment 1) first conducted a demonstration experiment using the six lists that had produced the most intrusions in Deese's (1959b) experiment: the 12 words most strongly associated with each of the words *chair*, *mountain*, *needle*, *rough*, *sleep*, and *sweet*. The subjects were instructed to pay close attention to the words as they were presented and, following presentation of each list, to recall as many words as they could without guessing. Roediger and McDermott confirmed Deese's (1959b) results with these lists, with the subjects recalling the critical nonpresented word from which the list was constructed on 40% of the trials. In fact, probability of recall of the critical nonpresented words approximated the probability of recall of words that had been presented in the middle of the list. This is a remarkable finding, one that was replicated in Experiment 2 and has been generally confirmed in subsequent work (McDermott, 1996; Schacter, Verfaellie, & Pradere, 1996).

Following presentation of all six lists, Roediger and McDermott (1995) gave subjects a recognition test in which studied items, critical lures, and other lures were presented. The subjects rated their confidence on a 4-point scale as to whether the word had appeared on the list (4 = *sure old*, 3 = *probably old*, etc.). Collapsing across level of confidence for old responses, the probability with which subjects classified critical items as being old (.86) was no greater than it was for critical lures (.84). Furthermore, over half of the critical lures (.58) were assigned to the "sure old" category. Unrelated lures were called "old" with a probability of only .02.

Encouraged by these results, showing remarkably high levels of false recall and false recognition, Roediger and McDermott (1995) conducted a second experiment in which they created 24 lists (the original 6 and 18 new ones) and again tested free recall and recognition. The recognition test employed the remember/know procedure developed by Tulving (1985) and Gardiner (1988). Briefly, in this procedure, subjects are told to classify each item judged old (or studied) as to whether they can recollect some specific aspect of its moment of occurrence in the list (a *remember* response) or know it was on the list but cannot remember the moment it was presented (a *know* response).

False recall of the critical item in Roediger and McDermott's (1995) Experiment 2 was 55%, even higher than it was in Experiment 1. In addition, the false alarm rate to the critical items (.76) was actually slightly greater than the hit rate for studied items (.72). Furthermore, the subjects were just as likely to report that they remembered some specific aspect about the occurrence of the critical nonpresented word in the study list (.48) as they were for studied words (.49). (These values are collapsed across two conditions in the experiment.)

In short, Roediger and McDermott's (1995) procedure represents a straightforward and powerful technique to elicit and to study false memories in a standard list-learning paradigm. As such, it offers a robust method to study memory illusions (Roediger, 1996). People recall, recognize, and remember events that did not occur. Furthermore, later research has shown that subjects will make an attribution about which of two experimenters, one male and one female, spoke the illusory word in a videotape presentation of the lists (Payne, Elie, Blackwell, & Neuschatz, 1996) and will attribute other characteristics to the illusory memories, albeit sometimes not at the same levels as for veridical memories (see Mather, Henkel, & Johnson, 1997; Norman & Schacter, 1997).

Given the wide interest in the topic of false memory and in using this technique to study it, we thought it would be helpful to obtain some normative data on lists that can be used to produce false memories. Deese (1959b) used 36 lists from Russell and Jenkins's (1954) word-association norms. Roediger and McDermott (1995, Experiment 1) used 6 of these lists in their first experiment and developed 18 more lists for their Experiment 2. McDermott (1995) developed 12 more lists. The current study employed all 36 lists. (Nineteen of these lists were based on ones used by Deese, 1959b). Here, we present normative data on false recall and false recognition for 36 lists developed by Roediger and McDermott (1995) and McDermott (1995).

METHOD

The general method followed that used by Roediger and McDermott (1995, Experiment 1), with the 24 lists from their Experiment 2 and the 12 additional lists developed by McDermott (1995).

Subjects

The subjects were 205 undergraduate students who were enrolled in introductory psychology at the University of Missouri–Columbia. Their participation was part of a course requirement.

Materials

Thirty-six lists were tested (see Appendix). Each list consisted of 15 *associates* of a *critical target*. For 24 of the lists (Roediger & McDermott, 1995), the 15 associates were generally the first 15 items in the Russell and Jenkins (1954) norms; some words were substituted when they seemed more likely to elicit the critical target or when a word appeared in another list or was a critical nonpresented target for another list. These 24 lists are marked with asterisks in the Appendix. The other 12 lists developed by McDermott (1995) were loosely constructed from the Russell and Jenkins norms, but the 15 associates selected were those that seemed most likely to produce a false memory. Again, some words not in those norms were used when they seemed more likely to lead to false memory or in order to avoid using words that were associates or critical targets in other lists.

The 36 lists were randomly divided into two sets of 18, and a random order of the lists within each set was constructed. The 18 lists were recorded onto audio tape in a male voice at the rate of 1 word every 2 sec. The items in each list were presented in the order shown in the Appendix (i.e., in order of strongest to weakest associative strength).

Recognition tests consisted of 108 words randomly ordered in 4 columns of 27 on a sheet of paper. The 108 words consisted of items from serial positions 1, 8, and 10 of each of the 18 studied lists (54 total items), the 18 critical targets, and 36 additional words not found in any of the 36 lists. The 36 additional words were selected to be unrelated to the 36 sets of words constituting the critical lists.

Procedure

The subjects were tested in one of four sessions, two devoted to each set of the lists; approximately 50 subjects participated in each session. In all, 95 subjects were tested with the first set of items, 110 with the second set. The subjects were told that this was an experiment on memory for lists of words, that they would hear several lists of words, and that at the end of each list (indicated when the experimenter pressed the pause button on the tape player), they should write on their answer sheets as many of the words as they could remember. As in Roediger and McDermott's (1995) study, the subjects were told to be reasonably sure that any word they wrote down was on the list. The subjects were given 2 min to recall each list. They recorded their responses in an answer booklet constructed of 18 sheets of paper, with the list number indicated at the top of each sheet. The subjects turned to the next answer sheet prior to presentation of each list.

After the subjects finished recalling the items from the last list, they turned their recall packets face down and were given the recognition test. They were asked to read each word carefully and to circle it if they thought it had been presented on any of the 18 lists. When all of the subjects had completed the recognition test, they were asked to write on the back of the recognition test what they thought the experiment was about.

Data Analyses

Each subject's responses were transcribed onto a coding sheet on which recall or recognition of an item was recorded. If, in answering the question about the purpose of the experiment, the subject mentioned that the words in the lists were all related to a common theme, they were judged to be at least somewhat aware of the purpose of the experiment. In Roediger and McDermott's (1995, Experiment 2) study, 1 subject indicated that the lists were designed to make her think of a nonpresented word; that subject was the only one to produce no false recalls. The criterion used in this study was much more liberal, in the hope of allowing a comparison of subjects aware of the structure of the lists and those who seemed not to be. After all the data were transcribed, they were double-entered into an electronic format. The two sets of data were then compared, and data entry errors were corrected before the results were analyzed.

RESULTS

The recall data are presented in Tables 1A and 1B, with the lists sorted according to the proportion of false recall of the critical target; Table 1A presents the 18 lists that produced the highest levels of false recall, and Table 1B presents the 18 that produced the lowest. Over all the lists, the range of false recall was fairly wide. Sixty-five percent of the subjects falsely recalled WINDOW when presented with the corresponding list, but only 10% falsely recalled KING. Roediger and McDermott (1995) reported that the mean level of false recall was close to that of the midlist items at the flat part of the serial position curve (Positions 4–11). Similarly, in the present study, recall of the actual list items followed the form of the standard serial position curve, and the probability of false recall (M = 40%) was fairly similar to the probability of recall at the middle of the serial position curve (again using Positions 4–11; M = 48%). For the top 18 lists (see Table 1A), however, false recall (M = 51%) was somewhat higher than recall for the midlist items (M = 45%). For the bottom 18 lists (see Table 1B), false recall (M = 29%) was lower than recall for the midlist items (M = 50%). The standard error of the mean was around 5% for recall of both the actual list items and the critical target within each list.

Veridical recall did not relate strongly to false recall; the Pearson product moment correlation across subjects of the proportion of the 270 list items correctly recalled with the

Table 1A
Proportion of Subjects Who Recalled an Item, as a Function of Serial Position and List, With Lists Ranked From Highest to Lowest by T for the Top 18 Lists

	Serial Position																
List	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	*M*	T
WINDOW	86	76	84	32	47	55	65	61	54	46	50	45	72	70	98	63	65
SLEEP	91	55	83	66	42	55	55	43	28	54	49	68	63	76	89	61	61
SMELL*	90	48	38	55	52	42	58	42	26	60	61	65	69	73	86	58	60
DOCTOR	62	85	57	50	37	77	43	33	66	38	44	86	36	47	88	57	60
SWEET	90	92	60	70	50	42	43	30	35	53	63	61	75	83	96	63	54
CHAIR	95	52	90	38	74	28	82	35	52	40	70	53	79	78	91	64	54
SMOKE	96	65	60	38	50	45	82	50	49	64	60	57	91	65	87	64	54
ROUGH	95	93	82	21	70	34	18	53	34	34	18	71	53	72	87	56	53
NEEDLE	74	59	72	60	33	32	57	51	70	20	39	80	77	83	94	60	52
ANGER	69	70	47	56	20	25	42	42	70	27	37	30	65	81	71	50	49
TRASH	93	73	66	50	18	40	27	22	33	20	29	68	86	88	93	54	49
SOFT	93	74	87	42	16	22	26	20	78	51	40	77	79	89	92	59	46
CITY	70	79	57	49	46	69	43	85	34	83	56	86	56	73	93	65	46
CUP	85	89	60	31	13	13	29	62	28	50	36	75	61	73	96	53	45
COLD	88	79	68	64	52	45	58	24	30	64	48	34	57	98	99	61	44
MOUNTAIN	87	80	49	52	31	64	43	26	38	61	58	71	68	76	95	60	42
SLOW	78	86	38	41	50	25	08	44	69	20	66	48	56	72	95	53	42
RIVER	78	86	68	90	45	28	33	54	43	61	63	76	64	80	97	64	42
M	84	75	65	50	41	41	45	43	47	47	49	64	67	77	92	59	51

Note—*M*, mean of 1–15; T, critical target. *See endnote 1.

Table 1B
Proportion of Subjects Who Recalled an Item, as a Function of Serial Position and List, With Lists Ranked From Highest to Lowest by T for the Bottom 18 Lists

	Serial Position																
List	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	*M*	T
SPIDER	95	82	61	60	27	91	44	69	44	38	64	33	52	74	92	62	37
FOOT	87	68	80	72	53	88	46	20	55	37	50	31	82	94	90	64	35
PEN	91	87	66	48	43	36	61	46	51	38	52	74	81	85	88	63	35
CAR	74	66	61	76	49	53	52	77	40	61	40	60	94	83	99	66	35
MUSIC	85	41	65	53	34	49	45	32	35	30	74	78	88	80	98	59	34
BLACK	75	77	72	63	47	72	37	28	62	45	36	47	54	82	98	60	34
RUBBER	71	71	41	42	65	33	26	26	33	22	80	56	49	88	93	53	32
GIRL	97	85	45	36	56	65	37	72	42	65	64	68	93	89	97	67	32
BREAD	93	46	47	44	53	51	29	20	64	34	37	38	85	78	100	55	31
FLAG	70	58	43	83	58	72	63	53	57	33	73	54	63	83	83	63	31
SHIRT	92	79	73	51	59	43	59	72	48	46	39	57	57	89	95	64	27
HIGH	77	83	42	40	45	50	40	56	32	38	66	50	78	78	95	58	26
ARMY	95	70	71	62	60	33	38	72	28	63	34	75	63	65	92	61	25
MAN	76	67	67	36	27	33	44	60	27	61	18	76	75	89	85	56	24
THIEF	84	80	40	30	53	85	63	66	51	35	38	33	83	77	94	61	23
LION	87	83	46	62	42	47	54	51	37	54	51	68	83	90	94	63	23
FRUIT	94	83	89	79	59	22	34	74	75	73	46	68	87	87	95	71	20
KING	97	84	70	72	77	55	27	37	39	43	64	74	53	82	96	65	10
M	86	73	60	56	50	54	44	52	46	45	51	58	73	83	94	62	29
Mean of Tables 1A and 1B	85	74	62	53	46	48	45	47	46	46	50	61	70	80	93	60	40

Note—*M*, mean of 1–15; T, critical target.

proportion of the 18 critical targets falsely recalled was $r = -.20$ ($n = 205$, $p < .01$). The correlation across lists was also modest; that is, across the 36 lists, false recall correlated $r = -.25$ ($n = 36$, $p > .10$) with veridical recall.

The level of false recall was very stable across different subgroups of subjects. Awareness of the structure of the lists seems not to have moderated performance. The correlation between proportion of false recalls for a given item in aware subjects ($n = 161$) with the corresponding proportion in unaware subjects ($n = 44$) was +.72 ($n = 36$, $p < .001$). Also, a split-half correlation was calculated by arbitrarily dividing the subjects into two groups, calculating the means for each list, and correlating the levels of false recall across lists. That correlation, too, was fairly high, with $r = +.80$ ($n = 36$, $p < .001$). Deese (1959b) reported levels of false recall for 19 of the 36 lists we used; the correlation between his level of false recall and ours for those lists was +.57 ($n = 19$, $p < .05$). Overall, the estimates of recall performance shown in Tables 1A–1B appear to be quite reliable.

The recognition data are presented in Tables 2A and 2B, with the lists sorted according to the proportion of false recognition of the critical target; Table 2A presents the 18 lists most likely to produce false recognition of the critical lure, and Table 2B, the 18 lists least likely to do so. The range here was remarkably wide, too. Eighty-four percent of the subjects falsely recognized WINDOW, but only 27% falsely recognized KING. The mean level of false recognition was 66%. The overall rate of false recognition of the critical target was about the same as the rate of correct recognition of the items from serial positions 8 ($M = 63\%$) and 10 ($M = 65\%$) but less than the rate of correct recognition of the item in Serial Position 1 ($M = 89\%$). For the top 18 lists (see Table 2A), however, false recognition ($M = 77\%$) was about midway between recognition of the items in Serial Position 1 ($M = 88\%$) and those in Serial Position 8 ($M = 60\%$) or 10 ($M = 66\%$). For the bottom 18 lists (see Table 2B), false recognition ($M = 55\%$) was lower than recognition for both the item in Serial Position 1 ($M = 91\%$) and the items in Serial Positions 8 ($M = 66\%$) and 10 ($M = 64\%$).

Table 2A
Proportion of Subjects Who Judged an Item to be Old on the Recognition Test, as a Function of Serial Position and List, With Lists Ranked From Highest to Lowest by T for the Top 18 Lists

	Serial Position				
List	1	8	10	*M*	T
WINDOW	87	63	65	72	84
SMELL*	93	64	90	82	84
COLD	92	66	78	79	84
ROUGH	96	71	47	71	83
CUP	91	66	73	77	82
SOFT	88	30	71	63	81
SLEEP	93	58	69	73	80
ANGER	81	63	64	69	79
SWEET	93	50	62	68	78
TRASH	96	62	50	69	78
CHAIR	90	62	71	74	74
SMOKE	97	68	86	84	73
HIGH	88	68	76	77	72
DOCTOR	69	68	68	68	71
THIEF	86	83	45	71	70
MOUNTAIN	90	40	76	69	69
SLOW	80	56	31	56	69
MUSIC	73	42	63	59	69
M	88	60	66	71	77

Note—*M*, mean of 1, 8, and 10; T, critical target. *See endnote 1.

Table 2B
Proportion of Subjects Who Judged an Item to be Old on the Recognition Test, as a Function of Serial Position and List, With Lists Ranked From Highest to Lowest by T for the Bottom 18 Lists

	Serial Position				
List	1	8	10	*M*	T
NEEDLE	90	78	46	71	68
RIVER	83	67	77	76	67
RUBBER	100	46	45	64	67
CITY	81	86	95	87	64
BREAD	93	44*	60	51	64
FOOT	87	32	59	59	62
MAN	89	67	83	80	61
FLAG	94	45	46	62	60
SPIDER	94	88	31	71	58
GIRL	89	89	85	88	58
PEN	91	67	46	68	57
SHIRT	95	92	66	84	54
ARMY	96	86	74	85	53
BLACK	77	44	66	62	49
FRUIT	96	76	85	86	45
CAR	84	91	75	83	42
LION	91	67	75	78	33
KING	99	68	40	69	27
M	91	66	64	74	55
Mean of Tables 2A and 2B	89	63	65	72	66

Note—*M*, Mean of 1, 8, and 10; T, critical target. *See endnote 2.

Similar to the recall results, veridical recognition did not relate strongly to false recognition; cumulating over lists, the correlation across subjects of the proportion of the 54 list items correctly recognized with the proportion of the 18 critical targets falsely recognized was +.15 (n = 205, $p < .05$). The correlation across lists was also small ($r = -.14$; $n = 36$, $p > .10$).

Like the recall data, the recognition results were very stable across different subgroups of subjects. The correlation between proportion of false recognition for a given item in aware subjects with the corresponding proportion in unaware subjects was +.79 (n = 36, $p < .001$), which was close to the split-half correlation of +.85 ($n = 36$, $p < .001$).

Finally, false recall and false recognition were strongly related. Across lists, the correlation was +.77 ($n = 36$, $p < .001$); across subjects, it was +.65 ($n = 205$, $p < .001$). Recall and recognition were not quite so strongly related for veridical memory, with correlations of +.52 (n = 36, $p < .001$) and +.49 (n = 205, $p < .001$) across lists and subjects, respectively. Of course, the relation between recall and recognition (both veridical and false) is likely to be at least partly due to the fact that recall preceded recognition in our design.

DISCUSSION

The present results provide a wealth of information about the effectiveness of the lists developed by Deese (1959b), Roediger and McDermott (1995), and McDermott (1995) in creating false recall and false recognition. This Deese–Roediger–McDermott paradigm, which has been dubbed the DRM (pronounced "dream") paradigm by Endel Tulving (see Roediger, McDermott, & Robinson, 1998), is being widely used by cognitive psychologists to study the illusion of remembering an event that never happened. Although many phenomena of false recognition have been studied, the DRM paradigm permits three relatively distinctive features in that (1) free recall can be used, (2) subjects are warned not to guess, and (3) testing occurs immediately after list presentation. Even when subjects are told the nature of the illusion and warned about producing the critical item, they still do so (Gallo, Roberts, & Seamon, 1997; McDermott & Roediger, 1998). Under certain conditions, the illusion can be remarkably powerful. For example, when McDermott (1996, Experiment 2) presented three lists blocked together for immediate free recall, subjects recalled 37% of the 45 list items correctly, but recalled the three critical items 57% of the time. The subjects were far more likely to recall the words that were not on the list than those that were.

Because we now know which lists create the effect most strongly, the norms we provide here permit the possibility of answering new questions about this phenomenon. Tables 1A and 2A provide the 18 lists that create the highest levels of false recall and false recognition, respectively. Tables 1B and 2B provide 18 lists that provide moderate amounts of false recall and false recognition. Of course, there is a high correlation between those lists providing false recall and false recognition, at least in our design, in which recognition always followed (and was therefore possibly influenced by) recall. Among the questions we are currently asking, using these norms, are the following: What is the role of associative processes in creating false memories in this and related paradigms (see Roediger et al., 1998)? Is there some characteristic or set of characteristics of the lists that determines their tendency to elicit false recall and false recognition? Can the lists that produce very high levels of false recall and false recognition be manipulated so as to dampen the phenomenon, by changing word order or by substituting a few new words for those in the current set? Correspondingly, can the lists that are less effective in producing false recall and false recognition be altered to be more effective in this regard? Can subjects (or for that matter, cognitive psychologists) judge (without our norms) which lists are more or less effective? We (McDermott, 1995; Roediger & McDermott, 1995) created the lists to produce high levels of false recall and false recognition and are ourselves mystified by the wide variability in their effectiveness and by which ones are and which ones are not effective. These are just a few of the questions that we think can be answered with the help of the norms provided here.

REFERENCES

BRUCE, D., & WINOGRAD, E. (1998). Remembering Deese's 1959 articles: The Zeitgeist, the sociology of science, and false memories. *Psychonomic Bulletin & Review*, **5**, 615-624.

Cofer, C. N. (1967). Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory* (pp. 181-214). New York: Wiley.
Deese, J. (1959a). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, **5**, 305-312.
Deese, J. (1959b). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, **58**, 17-22.
Deese, J. (1965). *The structure of associations in language and thought*. Baltimore: Johns Hopkins Press.
Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, **4**, 271-276.
Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, **16**, 309-313.
Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, **25**, 826-837.
McDermott, K. B. (1995). [Associative lists that induce false memories]. Unpublished materials.
McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language*, **35**, 212-230.
McDermott, K. B., & Roediger, H. L., III (1998). False recognition of associates can be resistant to an explicit warning to subjects and an immediate recognition probe. *Journal of Memory & Language*, **39**, 508-520.
Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, **25**, 838-848.
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 & Language*, **35**, 261-285.
Roediger, H. L., III (1996). Memory illusions. *Journal of Memory & 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, & Cognition*, **21**, 803-814.
Roediger, H. L., III, McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187-246). Hove, U.K.: Psychological Press.
Roediger, H. L., III, & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition*, **13**, 1-7.
Russell, W. A., & Jenkins, J. J. (1954). *The complete Minnesota norms for responses to 100 words from the Kent–Rosanoff Word Association Test*. (Tech. Rep. No. 11, Contract N8 ONR 66216, Office of Naval Research). Minneapolis: 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 & Language*, **35**, 319-334.
Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, **26**, 1-12.

NOTES

1. The word SMELL was inadvertently included as Item 14 in the FOOT list (we have corrected this problem in the list presented in the Appendix). By chance, both the FOOT list and the SMELL lists were presented to the same group of subjects, the former as their 15th list and the latter as their 18th. Because subjects were instructed to recall only those words on the current list, this is not likely to have affected the level of false recall for SMELL very much. It may have had more of an effect on recognition of this critical lure, but note that the lists' relative standing among all lists is about the same for both recall and recognition.

2. The word FLOUR (Item 8 on the BREAD list) was inadvertently misspelled "four" on the recognition test. Only 1% of the subjects responded positively to this item. The value shown in the table is an estimate of recognition for FLOUR, using a regression equation (r = .82) based on data from the other 35 lists to predict recognition performance on Item 8 from recall performance on that item.

(*Continued on next page*)

APPENDIX
Critical Targets With List Items 1 to 15

*ANGER: mad, fear, hate, rage, temper, fury, ire, wrath, happy, fight, hatred, mean, calm, emotion, enrage

ARMY: Navy, soldier, United States, rifle, Air Force, draft, military, Marines, march, infantry, captain, war, uniform, pilot, combat

*BLACK: white, dark, cat, charred, night, funeral, color, grief, blue, death, ink, bottom, coal, brown, gray

*BREAD: butter, food, eat, sandwich, rye, jam, milk, flour, jelly, dough, crust, slice, wine, loaf, toast

CAR: truck, bus, train, automobile, vehicle, drive, jeep, Ford, race, keys, garage, highway, sedan, van, taxi

*CHAIR: table, sit, legs, seat, couch, desk, recliner, sofa, wood, cushion, swivel, stool, sitting, rocking, bench

CITY: town, crowded, state, capital, streets, subway, country, New York, village, metropolis, big, Chicago, suburb, county, urban

*COLD: hot, snow, warm, winter, ice, wet, frigid, chilly, heat, weather, freeze, air, shiver, Arctic, frost

CUP: mug, saucer, tea, measuring, coaster, lid, handle, coffee, straw, goblet, soup, stein, drink, plastic, sip

*DOCTOR: nurse, sick, lawyer, medicine, health, hospital, dentist, physician, ill, patient, office, stethoscope, surgeon, clinic, cure

FLAG: banner, American, symbol, stars, anthem, stripes, pole, wave, raised, national, checkered, emblem, sign, freedom, pendant

*FOOT: shoe, hand, toe, kick, sandals, soccer, yard, walk, ankle, arm, boot, inch, sock, knee, mouth

*FRUIT: apple, vegetable, orange, kiwi, citrus, ripe, pear, banana, berry, cherry, basket, juice, salad, bowl, cocktail

*GIRL: boy, dolls, female, young, dress, pretty, hair, niece, dance, beautiful, cute, date, aunt, daughter, sister

*HIGH: low, clouds, up, tall, tower, jump, above, building, noon, cliff, sky, over, airplane, dive, elevate

*KING: queen, England, crown, prince, George, dictator, palace, throne, chess, rule, subjects, monarch, royal, leader, reign

LION: tiger, circus, jungle, tamer, den, cub, Africa, mane, cage, feline, roar, fierce, bears, hunt, pride

*MAN: woman, husband, uncle, lady, mouse, male, father, strong, friend, beard, person, handsome, muscle, suit, old

*MOUNTAIN: hill, valley, climb, summit, top, molehill, peak, plain, glacier, goat, bike, climber, range, steep, ski

*MUSIC: note, sound, piano, sing, radio, band, melody, horn, concert, instrument, symphony, jazz, orchestra, art, rhythm

*NEEDLE: thread, pin, eye, sewing, sharp, point, prick, thimble, haystack, thorn, hurt, injection, syringe, cloth, knitting

PEN: pencil, write, fountain, leak, quill, felt, Bic, scribble, crayon, Cross, tip, marker, red, cap, letter

*RIVER: water, stream, lake, Mississippi, boat, tide, swim, flow, run, barge, creek, brook, fish, bridge, winding

*ROUGH: smooth, bumpy, road, tough, sandpaper, jagged, ready, coarse, uneven, riders, rugged, sand, boards, ground, gravel

RUBBER: elastic, bounce, gloves, tire, ball, eraser, springy, foam, galoshes, soles, latex, glue, flexible, resilient, stretch

SHIRT: blouse, sleeves, pants, tie, button, shorts, iron, polo, collar, vest, pocket, jersey, belt, linen, cuffs

*SLEEP: bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap, peace, yawn, drowsy

*SLOW: fast, lethargic, stop, listless, snail, cautious, delay, traffic, turtle, hesitant, speed, quick, sluggish, wait, molasses

SMELL: nose, breathe, sniff, aroma, hear, see, nostril, whiff, scent, reek, stench, fragrance, perfume, salts, rose

SMOKE: cigarette, puff, blaze, billows, pollution, ashes, cigar, chimney, fire, tobacco, stink, pipe, lungs, flames, stain

*SOFT: hard, light, pillow, plush, loud, cotton, fur, touch, fluffy, feather, furry, downy, kitten, skin, tender

*SPIDER: web, insect, bug, fright, fly, arachnid, crawl, tarantula, poison, bite, creepy, animal, ugly, feelers, small

*SWEET: sour, candy, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, heart, cake, tart, pie

*THIEF: steal, robber, crook, burglar, money, cop, bad, rob, jail, gun, villain, crime, bank, bandit, criminal

TRASH: garbage, waste, can, refuse, sewage, bag, junk, rubbish, sweep, scraps, pile, dump, landfill, debris, litter

*WINDOW: door, glass, pane, shade, ledge, sill, house, open, curtain, frame, view, breeze, sash, screen, shutter

*Lists used by Roediger and McDermott (1995, Experiment 2). The other lists were developed by McDermott (1995).

(Manuscript received October 6, 1997;
revision accepted for publication March 9, 1998.)