

Factors that determine false recall: A multiple regression analysis

HENRY L. ROEDIGER III, JASON M. WATSON, KATHLEEN B. McDERMOTT,
and DAVID A. GALLO
Washington University, St. Louis, Missouri

In the Deese–Roediger–McDermott (DRM) paradigm, subjects study lists of words that are designed to elicit the recall of an associatively related critical item. The 55 lists we have developed provide levels of false recall ranging from .01 to .65, and understanding this variability should provide a key to understanding this memory illusion. Using a simultaneous multiple regression analysis, we assessed the contribution of seven factors in creating false recall of critical items in the DRM paradigm. This analysis accounted for approximately 68% of the variance in false recall, with two main predictors: associative connections from the study words to the critical item ($r = +.73$; semipartial $r = +.60$) and recallability of the lists ($r = -.43$; semipartial $r = -.34$). Taken together, the variance in false recall captured by these predictors accounted for 84% of the variance that can be explained, given the reliability of the false recall measures ($r = .90$). Therefore, the results of this analysis strongly constrain theories of false memory in this paradigm, suggesting that at least two factors determine the propensity of DRM lists to elicit false recall. The results fit well within the theoretical framework postulating that both semantic activation of the critical item and strategic monitoring processes influence the probability of false recall and false recognition in this paradigm.

In 1894, in the first volume of *Psychological Review*, E. A. Kirkpatrick performed “an experimental study of memory,” in which he showed that items presented as visual objects are better remembered than are words and that imagery aids retention of verbal materials. Those features of his paper have been cited, off and on, throughout the years. However, toward the end of the paper, Kirkpatrick reported briefly the results of a side experiment. It is worth quoting part of the relevant paragraph to help set the stage for the present inquiry:

About a week previously in experimenting upon mental imagery I had pronounced to the students ten common words. Many of these were recalled and placed with the memory list. Again, it appears that when such words as “spool,” “thimble,” “knife,” were pronounced many students at once thought of “thread,” “needle”, “fork,” which are so frequently associated with them. The result was that

many gave those words as belonging to the list. This is an excellent illustration of how things suggested to a person by an experience may be honestly reported by him as part of the experience. (Kirkpatrick, 1894, p. 608)

Kirkpatrick’s informal observations have recently been confirmed by many investigators (e.g., Roediger & McDermott, 1995), and false recall and false recognition owing to associative processes have become central topics in the field (see Roediger, McDermott, & Robinson, 1998, for a review). Roediger and McDermott’s (1995) research built on prior work by Deese (1959b). Deese presented subjects with 36 lists of 12 associatively related words, in which the list words (e.g., *shoe, hand, toe, kick*, etc.) were all associates of a word not presented in the list (*foot*, in this case), with a free recall test given after each list. For some of the lists, Deese (1959b) showed that subjects intruded the word associated to the list items at relatively high rates. However, Deese was interested in how associations affected recall (e.g., Deese, 1959a), not in false recall per se, and his paper and technique were relatively neglected (see Bruce & Winograd, 1998). Many of the lists Deese created elicited no or very few intrusions, which served his needs perfectly well because he was interested in how associative factors would affect both veridical and false recall (Deese, 1959a, 1959b). It is perhaps for this reason that Deese never made much of his intrusion findings. Indeed, in his 1965 book, *The Structure of Associations in Language and Thought*, which summarized much of his important work published over the previous 10 years, the intrusion paper (Deese, 1959b) was not even cited.

This research was supported by a contract from the U.S. Office of Technical Services to H.L.R. and K.B.M., as well as NIMH Grant 1RO3 MH59034-01 to K.B.M. The authors thank Martha Storandt for advice regarding the regression analysis, Michael A. Stadler for providing normative data, Douglas Nelson for advice in collecting our norms, David A. Balota for advice on various aspects of this research, Randy Buckner for calling our attention to Kirkpatrick’s work, and Cathy McEvoy, Douglas Nelson, and Tamaku Komatsu for providing additional data. Michael Cortese, Patrick Dolan, Stephanie Lee, Elizabeth Marsh, Michelle Meade, Douglas Nelson, and Valerie Reyna provided helpful comments. Correspondence concerning this article may be sent to H. L. Roediger III, Department of Psychology, Box 1125, Washington University, One Brookings Drive, St. Louis, MO 63130-4899 (e-mail: roediger@artsci.wustl.edu).

Roediger and McDermott (1995) developed a paradigm for the creation and experimental investigation of false memories that was an adaptation of Deese's technique. Following a suggestion made by Endel Tulving, this procedure is now referred to as the DRM paradigm, an acronym for Deese, Roediger, and McDermott. The basic procedure is straightforward: Subjects hear lists of 15 words that are the strongest associates to a missing word or critical item in word association norms (Nelson, McEvoy, & Schreiber, 1999; Russell & Jenkins, 1954). For example, subjects might be presented with a list of words like *bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap, peace, yawn, drowsy*—all of which are associated to the (nonpresented) critical item *sleep*. Immediately following the presentation of a study list, subjects recall as many of the list words as possible, without respect to their order (i.e., single-trial free recall). In addition, subjects are instructed to be sure that each word they write down was indeed on the preceding list. Despite this warning against guessing, subjects recall these critical nonpresented items with about the same probability as items appearing in the middle of the list (i.e., excluding the primacy and recency effects). At the end of the experiment, a recognition test that covers many lists is usually given (often with metamemory judgments). Recognition of the critical lures typically equals or exceeds recognition of the studied words. These remarkably high levels of false recall and false recognition in the DRM paradigm have been widely replicated (e.g., Payne, Elie, Blackwell, & Neuschatz, 1996, and Roediger, McDermott, & Pisoni, 2001, among many others).

The DRM paradigm is quite different from many other popular techniques used to study false memories. Most demonstrations of false memory phenomena have used relatively naturalistic materials, such as stories or visual scenes, fairly complex techniques, such as the presentation of misleading information, or relatively long delays between the study of the material and its later testing. In fact, the underlying assumption (beginning at least with Bartlett's, 1932, influential pronouncements) seems to be that these conditions are necessary to create illusory memories, because the use of such materials as word lists encourages reliance on reproductive rather than reconstructive memory. However, the levels of false remembering with the DRM paradigm are among the most robust ever obtained in the experimental literature, even though they occur under conditions of immediate testing, with no misleading information, with lists of words, and with strong warnings against guessing and making errors. Therefore, Roediger and McDermott (1995) suggested that the distinction between reproductive and reconstructive modes of retention should be abandoned. In some sense, all recollection, even immediate recall and recognition of lists of words, is reconstructive.

Although the DRM paradigm is a relative newcomer to the investigation of false memories, considerable research has been conducted using it and variations on its

basic theme. The purpose of this paper is to consider an overlooked issue in the growing literature on false memories—namely, the characteristics of the materials used to produce the effect. In Roediger and McDermott's (1995) first experiment, they selected the 6 lists used by Deese (1959b) that produced the highest levels of intrusion in his experiment, and they replicated his effects. In their second experiment, they used a total of 24 false memory lists that included the 6 lists from their first experiment, along with 18 new lists. Some of the new lists had been used by Deese, whereas Roediger and McDermott (1995) created some of the others, using similar procedures.

One remarkable outcome that we have noticed repeatedly while working with these lists is the striking variability they exhibit in eliciting false recall and false recognition of the associatively related critical item. These differences occur despite the fact that all the lists are created in the same manner, to include, with some minor variation, the first 15 associates of a single word that is itself not presented. With few exceptions, later researchers have used Roediger and McDermott's (1995) original 24 lists to study various interesting issues, collapsing the resulting data across the lists, as did Roediger and McDermott (1995). However, few investigators have been concerned with the fundamental issue of identifying characteristics of lists that influence the effect. That is, why do some lists work so well in producing the false memory effect, whereas others do not, when all of the lists are constructed in the same general manner? In addition, examination of the lists often reveals no obvious reasons for their variations in effectiveness at eliciting false recall. For example, consider the lists associated with *bitter* (*sweet, sour, taste, chocolate, rice, cold, lemon, angry, hard, mad, acid, almonds, herbs, grape, fruit*) and *sweet* (*sour, candy, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, heart, cake, tart, pie*). Despite the fact that both words refer to concepts in the same domain, represent the same part of speech, and have overlapping words in the lists, they differ dramatically in the effectiveness with which they elicit false recall. The mean probability of false recall of *bitter* when the relevant list was presented was only .01, whereas the mean false recall of *sweet* following the relevant list was .54.

Stadler, Roediger, and McDermott (1999) obtained norms on 36 DRM lists, the 24 used by Roediger and McDermott (1995) and 12 other lists. False recall ranged from greater than 60% intrusion rates to lists structured around the nonpresented words *window, sleep, smell, and doctor*, to 20% and 10%, respectively, for the lists structured around *fruit* and *king*. False recognition showed similarly wide variations. Although these lists were originally constructed to yield high levels of false responding, some did not. The purpose of the present study was to attempt to identify those characteristics of the lists that cause their differential efficacy in producing false recall and false recognition, through a simultaneous multiple regression analysis. By identifying

these factors, we should gain evidence bearing on the mechanisms that underlie the creation of false memories in the DRM paradigm. Before turning to this analysis, we review leads provided by prior research in uncovering the factors at work.

Deese's (1959a, 1959b) early work produced two important clues about factors that affect both veridical and false recall. The two factors he identified are associative connections among the study list words themselves and associative connections between the study list words and the critical item. Deese (1959a) provided subjects with lists that varied in what he termed *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. 235). He showed that this measure correlated highly ($r = +.88$) with the total number of words recalled from the lists—the stronger the associative bonds between items in the list, the better that list was recalled. More important for present purposes, interitem associative strength correlated negatively with the number of extralist intrusions ($r = -.48$). However, the stronger the associative bonds among list items, the more likely the subjects were to produce the same intrusion ($r = +.55$). Therefore, associations among list items increased recall of list items and decreased overall intrusions; yet, when intrusions did occur for strongly interconnected lists, they tended to be the same item across subjects.

Deese (1959b) also examined the influence of associative connections from the list words to the critical item on the false recall of the critical item. Subjects in this experiment heard 36 lists, each of which consisted of the 12 highest associates to the nonpresented critical item. False recall with these lists ranged from 42% for the critical item *needle* to 0% for the critical item *butterfly*. Deese (1959b) found that one predictor could account for the majority of the variance in the recall of critical items across his 36 lists. This variable was the mean backward associative strength (BAS), or the average tendency for words in the study list to elicit the critical item on a free association test. In fact, Deese (1959b) reported that the correlation between a list's mean BAS and the probability of recall of its associated critical item was $+ .87$. Roediger and McDermott (1995) argued that this outcome supports the notion that encoding factors are partly responsible for the effect. That is, lists high in BAS might activate the critical item, consciously or unconsciously, as the list items are presented. Subjects are confronted with a reality-monitoring dilemma (Johnson & Raye, 1981) during the test: "Did I hear the critical item, or did I only think of it while listening to the list items?"

Other data support this interpretation of the production of false memories in the DRM paradigm. Subjects claim to remember the occurrence of the critical items in the study list, which may mean that the item was consciously aroused during encoding (Roediger & McDermott, 1995) or, at least, unconsciously activated through spreading activation (Roediger, Balota, & Watson, 2001).

In addition, subjects claim to remember which of two speakers said the nonpresented critical item when the lists are presented in both a male and a female voice (Payne et al., 1996; Roediger, McDermott, & Pisoni, 2001). Similarly, when some lists are presented auditorily and others are presented visually, subjects attribute the illusory item as having occurred in the modality in which the list had been presented to the same extent as they make attributions for list items (Gallo, McDermott, Percer, & Roediger, 2001). Furthermore, subjects have been shown to report having rehearsed the critical item during the presentation of the study list (Mather, Henkel, & Johnson, 1997). Finally, McDermott (1997) showed that presentation of the lists yields priming of the nonpresented critical item on implicit memory tests of word-stem and word-fragment completion, albeit at a lesser magnitude than would be the case if the critical items had actually been presented. Because priming on these tests usually depends on lexical activation (Weldon, 1991), McDermott (1997) argued that her data were consistent with the notion that the presentation of DRM lists produces lexical activation of the critical item, at least for some lists.

Of course, all memory phenomena depend on the interaction of encoding and retrieval processes (Tulving, 1974, 1979), and we do not intend to suggest that retrieval processes are unimportant. In fact, Roediger and McDermott (1995) argued that the consideration of retrieval and decision processes are crucial to understanding this memory illusion and others. They cited Jacoby's attributional approach (e.g., Jacoby, Kelley, & Dywan, 1989) and Johnson's source-monitoring theory (e.g., Johnson, Hashtroudi, & Lindsay, 1993) as promising frameworks to guide this arena of inquiry. Israel and Schacter (1997) and Smith and Hunt (1998) have both argued that the distinctiveness of studied items at retrieval can explain why some manipulations reduce false recall and false recognition in this paradigm. Nevertheless, given the strong activation of the critical item by its semantic associates at encoding, list context clearly plays an important role in eliciting false memories in the DRM paradigm.

To our knowledge, McEvoy, Nelson, and Komatsu (1999) are the only other contemporary researchers who have attempted to determine the list-based factors that underlie the creation of false memories with associative lists. In their experiments, McEvoy et al. manipulated the two factors that Deese's (1959a, 1959b) earlier work had implicated: interitem associative strength (which they call connectivity) and the mean BAS of the list words to the critical item. Consistent with Deese's (1959a) research, they found that lists with high interitem associative strength yielded higher recall of list words and lower false recall of critical items, relative to lists with lower interitem associative strength. In addition, and also consistent with Deese (1959b), they found that lists with high mean BAS yielded more false recall than did lists with low mean BAS. They concluded that although con-

nectivity and mean BAS had opposite influences on the recall of critical items, both variables played an important role in creating false memories in the DRM paradigm.

In sum, both Deese (1959a, 1959b) and McEvoy et al. (1999) argued that at least two factors are important in predicting veridical and false recall. Associative connections from study words to the critical item increase the probability of false memories, whereas associative connections among study words decrease the probability of false memories. However, other factors besides mean BAS and connectivity may underlie the recall of critical items in the DRM paradigm. For example, consider the performance on the *king* list (i.e., *queen, England, crown, prince, George, dictator, palace, throne, chess, rule, subjects, monarch, royal, leader, reign*). Because it has a relatively high mean BAS, it would be predicted to have a high level of false recall. However, this list elicited the lowest level of recall in the entire set of 36 DRM lists normed by Stadler, Roediger, and McDermott (1999).

Gallo and Roediger (2001, Experiment 1) attempted to produce lists that were constructed in the same general way as those of Roediger and McDermott (1995) but that would, like the *king* list, produce low levels of false recall and false recognition. However, unlike the *king* list, the lists created by Gallo and Roediger all had low BAS. Of the 19 new lists they constructed, many had very low levels of false recall and false recognition. However, several of the lists (e.g., *wish, justice*) produced relatively high levels of false recall and false recognition, much like the lists developed by Roediger and McDermott (1995). Clearly, BAS is not the entire reason for the wide differences in the effectiveness of the lists in creating false memories. The aim of the present study is to gain further information on the characteristics of lists and critical items that affect false recall and false recognition.

OVERVIEW OF THE PRESENT STUDY

In the present study, we used a simultaneous multiple regression analysis to identify the factors predicting recall of critical items in the DRM paradigm. We considered factors that pertain to the critical item itself, as well as factors characterizing the study lists. Specifically, with respect to the critical items, we used the following three variables: word length, word frequency, and concreteness. On the basis of an examination of the Stadler et al. (1999) norms and Deese's (1959b) original findings, we hypothesized that factors making a critical item distinct might produce low levels of false recall. For example, *butterfly* is a long, low-frequency critical item. Interestingly, it was never recalled in Deese's (1959b) experiment, and it was recalled by only 1% of Gallo and Roediger's (2001) subjects.

We also assessed the ability of several variables related to the study words to predict the recall of critical items in the DRM paradigm. In particular, we used the following four variables: an index of the strength of the associative connections from the critical item to the

study words, or forward associative strength (FAS; i.e., associations from *sleep* to *bed*, to *rest*, to *awake*, etc.); an index of the strength of the associative connections from the study words to the critical item, or BAS (associations from *bed*, *rest*, . . . to *sleep*); the average number of associative connections among the study words (interitem associative strength, or connectivity); and the average probability of recall of the studied words within each of the false memory lists (veridical recall). On the basis of prior research by Deese (1959a, 1959b) and McEvoy et al. (1999), we hypothesized that BAS and connectivity would be the best predictors of the differences observed in false recall among critical items. More specifically, recall of critical items should increase as the strength of associative connections from the study words to the critical item increases; however, recall of critical items should decrease as the average number of associative connections among study words increases. (As will be shown, in our data set, only the first of these two predictions was fulfilled.) In sum, the goal of the present research was to identify the factors that underlie the creation of false memories in the DRM paradigm by explaining the wide variability in the effectiveness of our 55 lists to elicit false recall. The multiple regression analysis that led to our answers is explained more fully in the Method and Results sections.

METHOD

Materials

In the present study, we obtained data for 55 DRM lists, which included 24 lists from Roediger and McDermott (1995), 12 others from the Stadler et al. (1999) norms, and 19 lists from the Gallo and Roediger (2001) norms. The study words and the associated critical item for each list are presented in Appendix A, along with the BAS and FAS of each item in relation to the critical item. The words within lists are presented in descending order with respect to their strength of association to the critical item. That is, the first list item is typically the item most often generated as an associate to the critical item. Both veridical and false recall across our 55 lists are quite stable. The split-half correlation of veridical recall for the 55 lists is +.86, and the split-half correlation for the lists in eliciting false recall is +.90. Therefore, we could potentially explain 81% of the variance in false recall.

Variables

As was described in the introduction, the variables we used can be considered as belonging to two categories: properties of the critical item itself and properties of the word lists. A complete listing of the values of all seven variables (as well as levels of false recall and false recognition) for all 55 lists appears in Appendix B.

Critical item variables. With respect to the critical items, we used the following three variables: word length, log frequency, and concreteness. We consider each below.

1. *Word length* was defined by the number of letters in each critical item and represents the relative distinctiveness of each word form (longer words being more distinct). For example, *butterfly* has nine letters, whereas *sleep* has only five letters. We also obtained a subjective measure of *orthographic distinctiveness*¹ and an objective measure of *orthographic neighborhood size*,² but because of the high degree of multicollinearity among these three variables,³ we included only word length of the critical item in the multiple regression analyses.

2. The *raw frequency* of each critical item was obtained from the Kučera and Francis (1967) norms and is the number of instances found in print per million words. For example, the raw frequency of the critical item *sleep* is 65 and the raw frequency of the critical item *butterfly* is 2. To correct for skewness in the frequency distribution, for each critical item, raw frequencies were transformed into *log frequencies*, using the following formula: $\log(.5 + \text{raw frequency})$. For example, the log frequency of *sleep* is 1.82 [i.e., $\log(65.5)$], whereas the log frequency of *butterfly* is .40 [i.e., $\log(2.5)$].

3. *Concreteness* ratings for each critical item were obtained from the word association norms of Nelson et al. (1999), which were, in turn, obtained in part from the Paivio, Yuille, and Madigan (1968) and the Toglia and Battig (1978) norms. These ratings are obtained on a scale of 1–7, with 1 being the *least concrete* and 7 the *most concrete*. To carry through with our examples, *sleep* is slightly above the midpoint in concreteness (with a rating of 4.74), and *butterfly* is more concrete (with a rating of 5.91).

List variables. With respect to the study lists, we used the following four variables: FAS, BAS, interitem associative strength, and veridical recall. We consider each below.

1. *Forward associative strength.* Indices of the strength of associative connections from the critical item to the study words were obtained from the word association norms of Nelson et al. (1999). Specifically, for each critical item, we obtained the probability that it elicited each word in its corresponding study list as an associate, a measure Deese (1959b) referred to as the FAS. For example, as is shown in the third column of Appendix A, for the critical item *sleep* we identified the strength of its connections to *bed*, *rest*, *awake*, *tired*, and the remaining study words. We then averaged these associations to obtain the mean FAS value for each list. In the case of *sleep*, the mean FAS value for its 15 study words is .047, whereas, in the case of *butterfly*, the mean FAS value is .033. Note that all 15 list items were given as a response to the critical item in the Russell and Jenkins (1954) norms (from which the lists were generated). However, owing to discrepancies between these norms and those of Nelson et al. (1999), the number of forward connections for each list could have theoretically ranged from 0 to 15.

2. *Backward associative strength.* An index of the strength of associative connections from the study words to the critical item was obtained, in large part, from the Nelson et al. (1999) word association norms. However, for some of the 55 DRM lists, there were a few list words that were missing from these norms. For these words, we collected our own norms, using the procedures outlined by Nelson and colleagues. Briefly, we gave groups of over 75 undergraduates one of three sheets of 80 to-be-normed words; the subjects were instructed to write next to each word the first word it brought to mind. In those cases in which more than one word from a particular study list needed to be normed, these words were placed on separate norming sheets or were separated by at least 30 items on the same norming sheet. At least 86 observations were obtained for each normed item, and 150 or more observations were obtained for an item in some cases.

With these two sets of norms, we took each study word and determined its connection strength to its corresponding critical item, the measure Deese (1959b) referred to as the BAS. For example, as is shown in the second column of Appendix A, we obtained for each study word (e.g., *bed*, *rest*, *awake*, *tired*, etc.) the probability that it elicited its critical item (e.g., *sleep*) as an associate. Thus, this measure examines connections in the direction opposite to that used to obtain the FAS measures. We then averaged these associations to obtain the mean BAS value for each list. For example, the mean BAS value for the *sleep* list is .431. By comparison, the mean BAS value for the *butterfly* list is .045.

3. *Interitem associative strength, or connectivity.* The average interitem associative strength (Deese, 1959a), or connectivity (McEvoy et al., 1999), of the study words was obtained for each list using two steps. First, following a procedure outlined by McEvoy

et al., we constructed a 15×15 matrix of study items for each list. In Appendix C we have illustrated the connectivity matrix for the *sleep* list. Second, using the Nelson et al. (1999) norms and, in some instances, our own word association norms, we identified the presence of a connection for all 225 combinations of study word pairs in each list's matrix, excluding the 15 self-connections, which were not determined (listed as ND). For each word pair, a nonzero connection strength was scored as a 1, and a zero connection strength was scored as a 0. In the case of *sleep*, a total of 27 nonzero connections exists among the study words (or an average of 1.80 connections per study word). In the case of *butterfly*, a total of 29 nonzero connections exists among the study words (or an average of 1.93 connections per study word). High mean connectivity scores indicate a greater number of associative relationships among the study words, whereas low mean connectivity scores indicate a lesser number of associative relationships among the study words.

4. *Veridical recall of the list.* One final predictor was the average probability of recall of the studied words, collapsing across serial positions 1–15, for each list. These data were obtained from Stadler et al. (1999) and Gallo and Roediger (2001). In the case of *sleep*, studied items were recalled at a .61 level, whereas, in the case of *butterfly*, studied items were recalled at a .70 level. As was noted earlier, the split-half correlation of the veridical recall measure was reasonably high at +.86.

Procedure

Taking the seven factors that were just described, along with the probability of false recall and the probability of false recognition for each of the 55 lists, we constructed the data matrix shown in Appendix B for our simultaneous multiple regression analysis.⁴ Each of the 55 lists used in the regression analysis is represented by its respective critical item, each of which is listed in alphabetical order in the first column of the matrix. The second and third columns of the matrix contain the probability of false recall and false recognition, respectively, of each critical item from the Stadler et al. (1999) and Gallo and Roediger (2001) norms. The next three columns contain properties of the critical item itself: length, log frequency, and concreteness. The next four columns contain properties of the list that was constructed for each critical item: mean FAS, mean BAS, the mean connectivity of the study words, and the average recall of the studied items. The last three columns contain the critical item predictors that were excluded from the multiple regression analyses: raw frequency, orthographic distinctiveness, and orthographic neighborhood size (Coltheart's *N*).

In multiple regression analysis the independent variables can be entered either simultaneously (in a single step) or hierarchically (with some variables entered into the analysis before others). In the former case, all independent variables are allowed to compete for shared variance in the dependent variable (e.g., false recall) in a single step, so that no single variable is given greater priority than is another. In the latter case, those independent variables entered in earlier steps are allowed to capture variance in the absence of competition from variables to be entered subsequently. In this way, in a hierarchical analysis, earlier variables are given greater priority than later ones, usually for theoretical reasons. For the purpose of the present analysis, there exists no definitive theoretical justification for ordering the entry of critical item variables and list variables, so we opted for the simultaneous entry of all variables in a single step.

RESULTS

We consider first the bivariate correlations between our seven factors and false recall, and then we turn to the simultaneous multiple regression analysis and its findings. The entire matrix of correlations (Pearson's *r* in all

Table 1
Correlation Matrix Among Eight Variables: False Recall, Three Critical Item Variables, and Four List Variables

	False Recall	Length	Log Frequency	Concreteness	FAS	BAS	Connectivity	Veridical Recall
False recall	1.00							
Length	-.37*	1.00						
Log frequency	.17	-.40*	1.00					
Concreteness	.15	-.13	-.28*	1.00				
FAS	.08	-.15	.01	.10	1.00			
BAS	.73*	-.41*	.25	.26	.00	1.00		
Connectivity	-.04	.15	-.21	.19	.15	.10	1.00	
Veridical recall	-.43*	.31*	-.10	.26	.09	-.10	.32*	1.00

Note—FAS, forward associative strength; BAS, backward associative strength. * $p < .05$

cases) among the seven predictors and false recall is shown in Table 1.

Bivariate Correlations

In describing the results, we report significant bivariate correlations between critical item variables and false recall, followed by significant correlations between list variables and false recall. To determine these values, each of the critical item and list variables were entered into a regression equation as the sole predictor of false recall. Only three of our seven variables were significantly correlated ($p < .05$) with false recall: the critical item variable of length and the list variables of BAS and recall of the studied items (Figures 1A and 1B, respectively).

Length. The length of the critical item and false recall were negatively correlated [$r = -.37$, $t(53) = -2.87$].

Backward associative strength. Mean BAS, or the strength of the connections between study items and their corresponding critical items, was positively correlated with false recall, $r = +.73$, $t(53) = 7.73$. As is shown in Figure 1A, the greater the strength of association between the list items and the critical item, the more likely is false recall. This outcome replicates Deese's (1959b) finding of a $+ .87$ correlation between BAS and false recall, which he obtained with a set of thirty-six 12-item lists. Although the $+ .73$ correlation between these two variables in our set of fifty-five 15-item lists is somewhat smaller than the $+ .87$ correlation obtained by Deese, BAS is still the strongest correlate of false recall in our entire data set.

Veridical recall. Recall of list items and false recall were negatively correlated [$r = -.43$, $t(53) = -3.46$]. As is shown in Figure 1B, increasing levels of accurate recall are associated with decreasing levels of false recall. This outcome is contrary to some ideas about why false recall occurs, as will be brought out in the Discussion section.

As is shown in Table 1, no other variables were correlated with false recall at the bivariate level, which is surprising in some cases. For example, in terms of critical item variables, one might have expected word frequency or concreteness to be correlated with false recall, but neither did. (Deese, 1959b, also failed to find a relation between word frequency and false recall.) In terms of list variables, one might have expected connectivity to correlate with false recall, but it did not. The lack of a rela-

tion between connectivity and false recall is noteworthy because Deese (1959a) and McEvoy et al. (1999) reported a significant relationship between these variables, so that lists with greater connectivity were less likely to elicit false recall of the critical item. In the Discussion section, we consider connectivity, or interitem associative strength, more fully. We turn now to the results of the simultaneous multiple regression analysis.

Simultaneous Multiple Regression Analysis

In the multiple regression analysis, the three variables based on the characteristics of the critical items themselves (i.e., length, log frequency, and concreteness) and the four variables based on the characteristics of the lists designed to elicit the critical items (i.e., FAS and BAS measures, interitem associative strength, or connectivity, and the level of recall of list items) were entered simultaneously into the regression equation to predict false recall. The results of the analysis are shown in Table 2, including beta weights (β), semipartial correlations (sr), and significance tests for each predictor. For significance tests, a $t(47)$ was used for each predictor, and the resultant t statistic is expressed as the absolute value—that is, $|t|$. In addition, Table 2 displays the results of the significance test for the multiple R^2 for the entire set of seven predictor variables.

As is shown in Table 2, the results of the regression analysis indicated that two list variables were significantly related to false recall. Specifically, BAS was the strongest predictor of false recall, ($\beta = +.70$, $sr = +.60$, $t = 7.17$). Therefore, consistent with our findings at the bivariate level, the greater the strength of association between study items and their corresponding criti-

Table 2
Summary of Multiple Regression Analysis on False Recall With Critical Item Variables and List Variables Used as Predictors

Variable	β	Semipartial r	t value	R^2
Length	+.07	.06	.68	.68*
Log frequency	-.01	-.01	.06	
Concreteness	+.07	.06	.70	
Forward associative strength	+.12	.12	1.40	
Backward associative strength	+.70	.60	7.17*	
Interitem associative strength	-.02	-.02	.25	
Veridical recall	-.40	-.34	4.11*	

Note— $R^2 = .68$, $F(7,47) = 14.06$, $MS_e = 0.013$, $p < .05$. * $p < .05$.

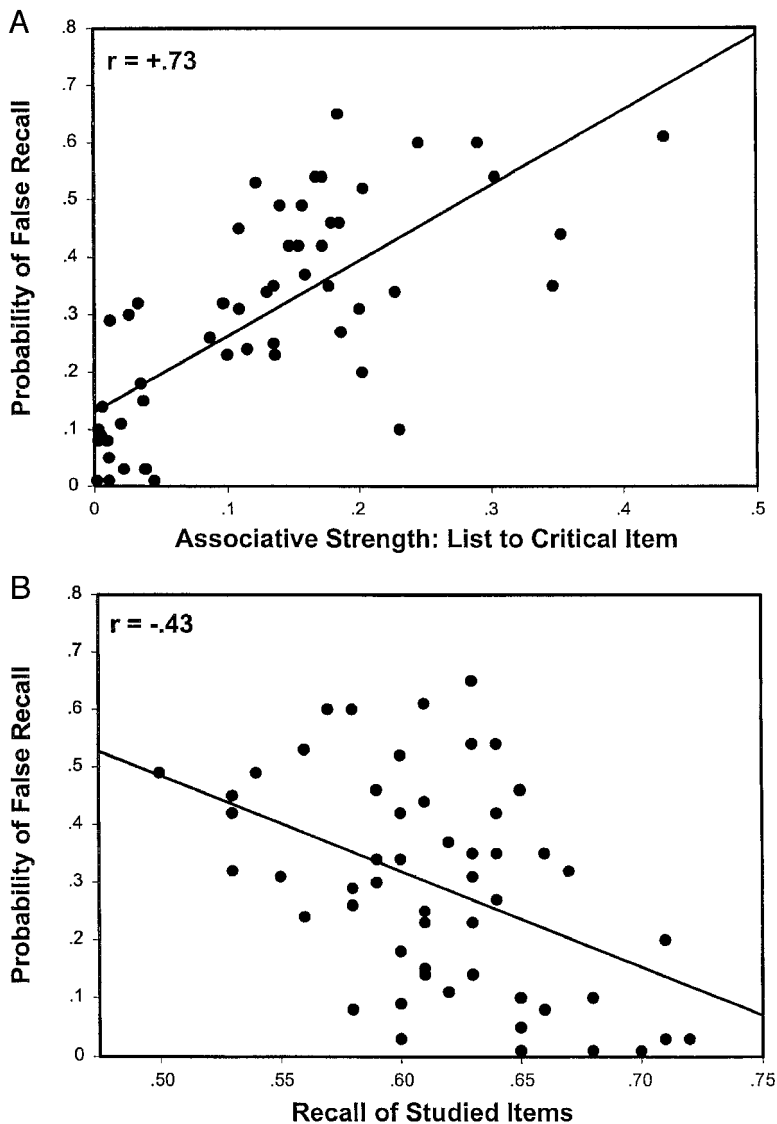


Figure 1. Probability of false recall of the critical item from 55 lists (A) as a function of associative strength of the list items to the critical item, or backward associative strength and (B) as a function of veridical recall of studied items from the lists.

cal item on a free association test, the more probable is false recall of the critical item. Veridical recall was also a strong predictor of false recall ($\beta = -.40$, $sr = -.34$, $t = 4.11$). As we found at the bivariate level, the more words recalled from a list, the less likely that the critical item for that list was falsely recalled. None of the remaining critical item or list variables, including length of the critical item, connectivity, and FAS, explained a unique portion of the total variance in false recall after the influence of other predictors had been residualized in the context of the regression equation. Therefore, BAS and recall of list items appeared to be the major contributors to the multiple R of $+.82$ and the associated multiple R^2 of $.68$ [$F(7,47) = 14.06$, $MS_e = 0.013$]. We further explore the implications of these effects in the Discussion section.

Factors Affecting False Recognition

In our analyses thus far, we have concentrated on false recall exclusively. However, the norms for our 55 lists (collected by Stadler et al., 1999, and Gallo & Roediger, 2001) also include values for false recognition. The recognition data in these norms were collected in a final recognition test, after lists had been studied and recalled. This confounding of recognition by prior recall almost certainly affected both veridical and false recognition. However, in experiments directly examining the effects of recall on recognition in this paradigm and others, the finding is that recall of a list may increase its recognition by a moderate amount (typically in the 4%–10% range; e.g., Roediger & McDermott, 1995; Roediger, McDermott, & Pisoni, 2001). In fact, some researchers have failed to observe any effect at all of recall

on recognition (Schacter, Verfaellie, & Pradere, 1996; see Roediger et al., 1998, for a discussion). We conducted the same regression analyses with our seven predictor variables on false recognition as we did on false recall. We summarize the basic results and conclusions as follows.

There were three significant bivariate correlations with false recognition, including length ($r = -.27$), BAS ($r = +.43$), and veridical recall ($r = -.52$). Overall, these correlations with false recognition were similar to those obtained with false recall. However, the positive correlation between BAS and false recognition was smaller than that in the recall data, and the negative correlation between veridical recall and false recognition was larger than that in the recall data. To maintain consistency with the previous regression analysis on false recall, we used veridical recall, rather than veridical recognition, in this analysis. In addition, recognition estimates were based on only three items/list and were probably affected by prior recall.

Turning to the results of the simultaneous multiple regression analysis on false recognition, BAS and veridical recall of list items appeared to be the major contributors to the multiple R of $+.69$ and the associated multiple R^2 of $.48$ [$F(7,47) = 6.11$, $MS_e = 0.02$]. Table 3 shows the summary of the regression analyses. Although the regression analysis accounted for less total variance in false recognition than in false recall ($.48$ as opposed to $.68$, respectively), the same factors seem to be responsible for predicting false recognition and false recall. Of course, because measures of recognition always followed recall, this similar pattern may reflect a carry-over effect of false recall on false recognition. Nonetheless, within the limits of this study, the factors responsible for false recognition across lists seem to be the same as those producing false recall.

DISCUSSION

To summarize the main findings, our results implicate two primary factors (of the seven variables we examined) as accounting for most of the variance in false recall across lists. The split-half correlation of our lists was $+.90$, so the amount of potentially explainable variance in our dependent variable of false recall is 81%. The fac-

tors in our multiple regression accounted for 68% of the total variance in false recall, or 84% of the explainable variance ($.68 \div .81 \times 100 = 84\%$). The analysis for false recognition revealed the same factors at work, although they accounted for less overall variance. The strongest factor was BAS, consistent with prior work by Deese (1959b) and McEvoy et al. (1999). The other factor was the number of items recalled accurately from the list, which is negatively correlated with false recall.

We turn next to the theoretical implications of our findings. We first consider the implications for the activation/monitoring framework that we have developed to explain false recall and false recognition in the DRM paradigm. We then consider the implications for other theories and finally turn to several further issues on which our results shed light.

Activation/Monitoring Theory

In the theoretical framework that we (e.g., McDermott & Watson, 2001; Roediger, Balota, & Watson, 2001; Roediger & McDermott, 1995, 2000) have used to explain variations in the probability of false recall and false recognition in the DRM paradigm, two sets of processes have been proposed. Specifically, we have proposed that processes occurring during both encoding and retrieval affect the probability of false recall and false recognition in the DRM paradigm. Two critical sets of processes involve activation and monitoring. Although there is a natural alignment of activation to encoding and monitoring to retrieval, we note that both processes can potentially occur both during encoding and during retrieval, as will be discussed below. The ideas we employ have many precursors, which we acknowledge below, although their particular arrangement here may be unique.

A critical distinction in understanding encoding processes and their interaction with retrieval processes is between item-specific and relational processing (Anderson, 1972; Hunt & Einstein, 1981; Hunt & McDaniel, 1993). Briefly, the notion is that emphasis can be given to the encoding of individual elements and their features (what Anderson called *node-tagging* and Hunt and his colleagues called *item-specific processing*) or to the encoding of relationships among elements (*pathway-tagging* or *relational processing*). Item-specific processing should

Table 3
Summary of Multiple Regression Analysis
on False Recognition With Critical Item Variables
and List Variables Used as Predictors

Variable	r	β	Semipartial r	t value	R^2
Length	-.27*	+.09	.07	0.63	.48*
Log frequency	.21	+.12	.09	0.89	
Concreteness	-.11	-.07	-.06	0.56	
Forward associative strength	.12	+.16	.16	1.47	
Backward associative strength	.43*	+.38	.32	3.02*	
Interitem associative strength	.03	+.17	.15	1.44	
Veridical recall	-.52*	-.55	-.46	4.38*	

Note— $R^2 = .48$, $F(7,47) = 6.11$, $MS_e = 0.02$, $p < .05$. * $p < .05$.

make individual elements more memorable and lead to more distinctive memories; such processing should help distinguish veridical memories from false memories, because list items would have features of presentation associated with them, whereas the critical item sometimes would not. (However, if the critical item is strongly activated during encoding, it may take on many of the features of list items.) Relational processing (extracting themes or schemas or constructing networks of associations) should increase the probability of false recall and false recognition by leading to inferences or associations of events that may not actually have occurred.

We assume that in encoding and interpreting experiences, people go beyond the information given in the stimulus (Bartlett, 1932, chap. 2; Bruner, 1957) and activate related information through inferential processing. Each person's experience is coded in terms of his or her own schemata (or associative networks or knowledge structures). When schemata are activated, a process of filling in of "missing" information may occur, whereby items strongly implied by the schema (but not actually present in the stimulus) may become activated (Bartlett, 1932). In encoding lists of words in the DRM paradigm, presentation of the list can create activation that spreads throughout the lexical/semantic system and can create implicit associative responses (Underwood, 1965). This activation may arouse concepts that are not presented but that are associated to concepts that are presented (e.g., Anderson & Pirolli, 1984; Collins & Loftus, 1975; Meyer & Schvaneveldt, 1971). The associative activation in this approach can arise from automatic processing or from more controlled, elaborative processing (Neely, 1977). In addition, the associative activation of concepts that were not presented may occur consciously (subjects may become aware of the concept of *sleep* and rehearse it during presentation of the list words *bed*, *rest*, *awake*, etc.), or the concept may be strongly activated but never come to conscious awareness. The debate over whether arousal of the critical item in the DRM paradigm is conscious or unconscious is interesting but difficult to decide empirically. Some evidence has implicated conscious arousal of the concepts (McDermott, 1997), whereas other evidence implicates unconscious activation (Seamon, Luo, & Gallo, 1998). The correct answer to the question of whether activation is conscious or unconscious is probably "both"—that is, the critical item probably sometimes comes consciously to mind and, even if it does not, relevant information may nevertheless be activated unconsciously.

The high correlation between BAS and false recall, as well as the strong weighting of BAS in the multiple regression analysis, strongly implicates activation of critical items as an important factor in explaining false recall and false recognition in the DRM paradigm. The more strongly associated list items are to critical items, the more activation would be expected to spread to the critical items. More colloquially, to the extent that the words on the list spark the associative connection to the critical

item, the more likely the critical item is to be falsely recalled. This activation factor would seem to be one main reason why "things suggested to a person by an experience may be honestly reported by him as part of the experience" (Kirkpatrick, 1894, p. 608). These associated lists probably all tend to elicit relational processing, but only those with strong BAS to the critical item lead to false recall.

The second set of factors proposed in activation/monitoring theory is concerned with the monitoring of memory accuracy. Monitoring of encoding processes can occur during the study phase of an experiment, especially under intentional learning conditions. In experiments in which subjects are warned about false memory phenomena and are told to pay close attention to lists and to remember only the presented items, and not the critical items, monitoring is enhanced (Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001; McDermott & Roediger, 1998). Monitoring of memories has typically been discussed with respect to retrieval, during which time a primary goal is to distinguish, from among the considerable information brought to consciousness, that which refers to perception of past events from that which does not, a process referred to as *reality monitoring* (Johnson & Raye, 1981). The experience of remembering events involves mentally traveling back in time and seeming to reexperience the events (Tulving, 1985). During retrieval, recovered memories that apparently carry with them details of having been experienced are generally attributed to be memories of actual past events, although sometimes this attribution may be in error (Jacoby et al., 1989).

Johnson et al. (1993) have described in detail the types of information that may be used to distinguish events that actually happened from ones that were only imagined, thought about, or read about. For example, events that actually occurred have more features from the external world associated with them (sights, sounds, etc.), whereas those that were internally generated are less likely to have these features, but more likely to have associated reflections and cognitive operations. In the DRM paradigm, the fact that the critical item has been strongly activated during encoding means that, during retrieval, the subject is likely to retrieve the critical item as a list member. To the extent that the critical items are strongly activated during encoding, they will take on some features similar to those of the list items; this is especially true if the critical items come consciously to mind during encoding and are rehearsed in the same manner as list items. To the extent that the features of the critical item are similar to those of actual list items, the probability of false recall or false recognition is increased. However, if the list items systematically differ from the critical item on some dimensions or features, retrieval heuristics may be brought into play that can reduce false recall and false recognition (Israel & Schacter, 1997; Schacter, Israel, & Racine, 1999). Considerable evidence supports this general source-/reality-monitoring frame-

work that we advocate here (e.g., Johnson et al., 1993; Schacter et al., 1999), as well as the attributional approach to remembering (Jacoby et al., 1989).

The other significant factor arising from our multiple regression analysis may operate during retrieval by aiding the monitoring processes that reduce false recall. The negative correlation between veridical recall of list items and false recall of the critical item may indicate that the better encoded list items are, the more easily they can be distinguished from the illusory critical item. Subjects may use item-specific information to aid recall of list items and suppress recall of the critical items, which would not carry as much item-specific information. When experimental manipulations increase veridical recall, either through longer study time at relatively slow rates (Gallo & Roediger, 2001; McDermott & Watson, 2001; Toglia, Neuschatz, & Goodwin, 1999) or through repeated study and test opportunities (McDermott, 1996; Schacter et al., 1999), false recall is often decreased. Therefore, the negative relation between veridical recall and false recall can be produced experimentally, as well as appearing in our correlation/regression analysis. However, veridical and false recall can be positively related as a function of other variables, too, as was shown by McDermott (1996), Toglia et al. (1999), and Thapar & McDermott (2001). In these cases, greater activation of the critical item from relational processing or from greater semantic processing is probably at work.

The fact that the length and/or distinctiveness of the critical item are negatively correlated with false recall can be similarly interpreted as the effect of monitoring during retrieval, because the greater distinctiveness of the critical item simplifies the monitoring process. For example, if words in the list are generally shorter and have more regular orthography than does the critical item, then during retrieval the critical items that are long and distinctive (e.g., *butterfly*, *cabbage*) should be more easily discriminated from list items, as indeed seems to be the case in the bivariate correlations. Interestingly, the length/distinctiveness factor played a role in determining false recall even though the lists in our norms were presented auditorily and the orthographic information, at least, should not be particularly accessible. In addition, presenting the DRM lists visually, rather than auditorily, reduces false recall and false recognition, at least on tests in which recognition items are presented visually (as opposed to auditorily) or recall is written rather than spoken (Gallo, McDermott, et al., 2001; Kellogg, 2001; Smith & Hunt, 1998). This modality difference has also been interpreted as arising from the retrieval of distinctive information operating during the test (Gallo, McDermott, et al., 2001; Smith & Hunt, 1998).

As was noted above, activation processes may occur during retrieval, as well as during encoding. The more list items a subject recalls, the more primed the critical item might become. However, the present analyses undercut the simplest form of this idea. If the recalling of list items activated or primed false recall, the simplest

prediction would be that there should be a positive correlation between veridical recall and false recall. Our data revealed the opposite pattern, showing a negative correlation between veridical and false recall. Marsh, McDermott, and Roediger (2001) reported other experiments that also call into question the idea that priming during the test plays a role in false recall and false recognition.

Balota et al. (1999) obtained further evidence supporting the activation/monitoring interpretation of false memories in the DRM paradigm. They compared performance of younger adults, healthy older adults, and older adults with Alzheimer's disease. The latter two groups have been shown to have intact activation processes, as measured by semantic priming paradigms (Balota & Duchek, 1988, 1991), yet they are deficient in source-monitoring processes, which is probably attributable to impaired frontal lobe functioning (see Balota, Dolan, & Duchek, 2000). Therefore, one might expect that whereas veridical recall would be worse in older adults and Alzheimer's patients (relative to younger adults), false recall and false recognition would occur at the same level or, perhaps, be increased in these groups. Indeed, Balota et al. (1999) reported exactly this pattern. A similar outcome has been obtained in other laboratories (e.g., Budson, Daffner, Desikan, & Schacter, 2000; Norman & Schacter, 1997; Tun, Wingfield, Rosen, & Blanchard, 1998). This pattern is just as predicted by the activation/monitoring framework.

In short, the present results and many others (see Roediger, Balota, & Watson, 2001; Roediger et al., 1998) fit comfortably within the activation/monitoring framework. In the remainder of our discussion, we will consider briefly the implications of our present results for other accounts of the DRM paradigm and for the role of connectivity (or interitem associative strength) in producing veridical and false recall.

Fuzzy-Trace Theory

Another leading theory that has been used to explain false memory phenomena is fuzzy-trace theory (e.g., Reyna & Brainerd, 1995). Payne et al. (1996) and Schacter et al. (1996), among many others, have used the basic theory to account for false recall and false recognition data within the DRM paradigm, and the theory has been extended in interesting ways to account for new variations on the paradigm (e.g., Brainerd, Wright, Reyna, & Mojardin, 2001). The fundamental assumption of fuzzy-trace theory is that experience leaves two types of memory traces; gist traces (which capture the meaning of experience, but without specific attributes) and verbatim traces (which represent specific attributes). False memories are attributed to gist processing: When the meaning of a recognition probe matches the gist of the encoded experience, false recognition occurs. The theory has been directed primarily to false recognition and not false recall.

In trying to explain the present results, one difficulty is that the operational definition of *gist* has never been specified; that is, within the DRM paradigm, given a list

of 15 related words, how does one specify the gist representation of the list and quantify how strong it is? One could use patterns of false recognition—to the extent that a concept is falsely recognized, it was represented as part of the gist representation. This definition is perfectly suitable for some purposes, but clearly explaining false recognition is not one of them, owing to the problem of circularity. To explain why some lists lead to high levels of false recall and false recognition (whereas other lists lead to low levels) within fuzzy-trace theory, one needs an independent specification of gist, which, to our knowledge, has not been provided in this context. Why does the *sweet* list have so much gist that it leads to false recall, whereas the *bitter* list does not? One plausible interpretation of gist is BAS: The more the list items are associated to the critical item, the more a gist representation of the critical item is created. This solution to a definition of gist works within the DRM paradigm, but, of course, the interpretation of how gist is created becomes isomorphic to activation processes within the activation/monitoring framework.

Interpretation of the other factor that was significant within our multiple regression analysis may be more straightforward within fuzzy-trace theory. The negative correlation between veridical recall and false recall (and false recognition) may be accounted for by increased reliance on verbatim traces. That is, if increased recall of list items is due to increased verbatim information, subjects may rely less heavily on gist information and, hence, show less false recall and false recognition during responding. Similarly, for critical items that are long and/or orthographically distinct, recall of list items that relies on these features may lead to suppression of false recall and false recognition of the critical items, because they may differ systematically from other list items. These factors postulated within the context of fuzzy-trace theory are quite similar to those within the activation/monitoring framework outlined above. At some level of generality, there are two sets of factors in each account, and both theories can explain many empirical results.

Shifting Criterion Theory

Miller and Wolford (1999) proposed that false recognition in the DRM paradigm could be explained by subjects' shifting their criteria differentially to list items and to critical items during a recognition test. (They extended their explanation to false recall, but somewhat tenuously, by proposing a generate/recognize theory.) They postulated that subjects encoded the structure of the list during its presentation and then, during the recognition test, responded with especially liberal criteria to critical items, but not to list items. The criterion shift is supposed to occur to the degree that the test item appears to be related to the list items. To support their argument, they reported that a measure of bias from signal detection theory was indeed more liberal for critical items than for list items. However, Wickens and Hirshman (2000) and Wixted and Stretch (2000) pointed out that,

within signal detection theory, differences in bias measures do not necessarily implicate a criterion shift. Rather, differences in measured bias could arise from differences among the underlying distributions. Wixted and Stretch argued that this latter interpretation was the more likely and that the criterion shift interpretation was untenable (also see Roediger & McDermott, 1999).

The present evidence also weighs heavily against a criterion shift account. Miller and Wolford (1999) appealed to subjects' metaknowledge of the structure of the list. Within the criterion shift account, we can assume that the likelihood that a critical item cues metaknowledge of the list depends on its FAS, or how well the critical item will cue the list. If so, FAS (the association of the critical item to the list items) should be the primary determinant of false recognition and false recall. However, the correlation between FAS and false recall in our data was negligible (.08), and the same was true in analyses of false recognition. BAS was actually the primary predictor of false recall, which implicates activation processes during encoding as being a primary cause of the effect. There would be no reason to expect BAS to have any impact, according to the idea that a simple criterion shift during the test causes false recall and false recognition. Of course, all our lists were relatively high and roughly equivalent in terms of FAS, because they were created from forward associations to the critical item, so restriction of range may account for the low correlation. Even so, the same problem exists for the shifting criterion theory to explain the results. If FAS is supposed to determine false recall and false recognition by this account and if all our lists were high and roughly equal on this feature, why were so many lists ineffective in eliciting false recognition and false recall? In short, either way, the shifting criterion theory cannot account for the results.

Of course, within the activation/monitoring framework, the monitoring processes occurring during retrieval could be conceived as conservative shifts in response criteria. However, in our theory, these monitoring processes determine the degree to which a given level of activation at retrieval will be judged sufficient to endorse a critical item as having been studied, thereby producing a false memory. In contrast, in Miller and Wolford's (1999) account, there was no activation of the critical item during encoding, and shifting criteria during the test were said to account entirely for the false recognition phenomenon (see Wixted & Stretch, 2000). In the activation/monitoring account, the memory illusion provoked by the DRM lists is assumed to be a subjectively powerful phenomenon that can be ameliorated (but rarely eliminated) when list items are made more distinctive, relative to critical items. Under conditions in which the subject is warned at test (but not before the encoding phase) to monitor their memories for the presence or absence of the critical item and is discouraged from making criterion shifts, the false memory effect is neither eliminated nor greatly reduced (Gallo, Roediger, & McDermott, 2001). Hicks and Marsh (2001) showed that under conditions in which

subjects are explicitly told to monitor their memories during the test (and are asked to judge whether each word had been externally presented or internally generated), false recognition in the DRM paradigm actually increases rather than decreases. In fact, the false recognition effect persists even when subjects are fully informed about the nature of the phenomenon before the study episode and are therefore encouraged to enhance monitoring processes during both encoding and retrieval in order to attempt to diminish or eliminate the effect (see Gallo et al., 1997; Gallo, Roediger, & McDermott, 2001; McDermott & Roediger, 1998). For these and other reasons, the notion that the DRM memory illusion is caused by a general shift of criteria can be eliminated as a potential explanation.

Interitem Associative Strength or Connectivity

Our data pose something of a puzzle about the role of interitem associative strength (Deese, 1959a) or connectivity (McEvoy et al., 1999) in false recall. As was discussed in the introduction, Deese (1959a) found a significant negative correlation between this construct and the probability of false recall. In a later paper, Deese (1961) obtained a weaker negative correlation that failed to reach conventional levels of significance, so the issue is somewhat in doubt from Deese's (1959a, 1961) research. McEvoy et al. created lists that varied in connectivity (interitem associative strength) and in BAS and manipulated these factors orthogonally. They replicated the finding of Deese (1959a); in that the density of interconnections of list words was negatively related to false recall. However, in our analysis of 55 lists (different from those used by McEvoy et al., 1999), we found no relation between connectivity and false recall ($r = -.04$) or between connectivity and false recognition ($r = .03$).

Further research will be necessary to determine the true state of affairs, although a case can be made that the lack of correlation we obtained is accurate. First, we had 55 lists that varied widely in their ability to elicit false recall, as compared with 24 lists used in McEvoy et al. (1999). Our lists also represented a relatively full range of connectivity, or interitem associative strength, relative to those of McEvoy et al. McEvoy et al. used lists that had a somewhat wider range of connectivity, because they used extreme values, whereas our larger set of lists provided a fuller body of the middle ranges of connectivity. Given our results, it seems at least possible that interitem associative strength (connectivity) is not a factor in determining whether lists of associative words lead to false recall, although the issue remains in doubt.

We did find that interitem associative strength, or connectivity, is positively and significantly correlated with veridical recall ($r = +.32$). This outcome is consistent with that of Deese (1959a) and McEvoy et al. (1999). The greater the associative bonds among list items, the greater is list recall. This outcome is also consistent with virtually any associative theory of recall (e.g., Anderson, 1972, 1983; Nelson, Schreiber, & McEvoy, 1992).

CONCLUSION

The present study has uncovered two primary factors that are significantly related to false recall and false recognition in the DRM paradigm. BAS is positively correlated with false recall, whereas veridical recall of the list items is negatively related to false recall. The activation/monitoring account provides a natural interpretation of these factors. Any viable theory posited to explain the occurrence of false memories in the DRM paradigm must explain the variability in false recall and false recognition across lists and the two primary factors responsible for this variability.

REFERENCES

- ANDERSON, J. R. (1972). FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 5, pp. 315-378). New York: Academic Press.
- ANDERSON, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- ANDERSON, J. R., & PIROLI, P. L. (1984). Spread of activation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 791-798.
- BALOTA, D. A., CORTESE, M. J., DUCHEK, J. M., ADAMS, D., ROEDIGER, H. L., III, McDERMOTT, K. B., & YERYS, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimers type. *Cognitive Neuropsychology*, **16**, 361-384.
- BALOTA, D. A., DOLAN, P. O., & DUCHEK, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395-409). Oxford: Oxford University Press.
- BALOTA, D. A., & DUCHEK, J. M. (1988). Age-related differences in lexical access, spreading activation, and simple pronunciation. *Psychology and Aging*, **3**, 84-93.
- BALOTA, D. A. & DUCHEK, J. (1991). Semantic priming effects, lexical repetition effects, and contextual disambiguation effects in healthy aged individuals and individuals with senile dementia of the Alzheimer type. *Brain & Language*, **40**, 181-201.
- BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- BRAINERD, C. J., WRIGHT, R., REYNA, V. F., & MOJARDIN, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 307-327.
- 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.
- BRUNER, J. S. (1957). On perceptual readiness. *Psychological Review*, **64**, 123-152.
- BUDSON, A. E., DAFFNER, K. R., DESIKAN, R., & SCHACTER, D. L. (2000). When false recognition is unopposed by true recognition: Gist-based memory distortion in Alzheimer's disease. *Neuropsychology*, **14**, 277-287.
- COLLINS, A. M., & LOFTUS, E. F. (1975). A spreading-activation theory of semantic memory. *Psychological Review*, **82**, 407-428.
- COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.
- DEESE, J. (1959a). Influence of interitem associative strength upon immediate free recall. *Psychological Reports*, **5**, 235-241.
- DEESE, J. (1959b). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, **58**, 17-22.
- DEESE, J. (1961). Associative structure and the serial reproduction experiment. *Journal of Abnormal & Social Psychology*, **63**, 95-100.
- DEESE, J. (1965). *The structure of associations in language and thought*. Baltimore: Johns Hopkins University Press.

- GALLO, D. A., McDERMOTT, K. B., PERCER, J. M., & ROEDIGER, H. L., III. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 339-353.
- 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.
- GALLO, D. A., & ROEDIGER, H. L., III (2001). *Variability among word lists in evoking associative memory illusions*. Manuscript submitted for publication.
- GALLO, D. A., ROEDIGER, H. L., III, & McDERMOTT, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, **8**, 579-586.
- HICKS, J. L., & MARSH, R. L. (2001). False recognition occurs more frequently during source identification than during old-new recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 375-383.
- HUNT, R. R., & EINSTEIN, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning & Verbal Behavior*, **19**, 497-514.
- HUNT, R. R., & McDANIEL, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory & Language*, **32**, 421-445.
- ISRAEL, L., & SCHACTER, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, **4**, 577-581.
- JACOBY, L. L., KELLEY, C. M., & DYWAN, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.
- JOHNSON, M. K., HASHIROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.
- KELLOGG, R. T. (2001). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 913-919.
- KIRKPATRICK, E. A. (1894). An experimental study of memory. *Psychological Review*, **1**, 602-609.
- KUČERA, H., & FRANCIS, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- MARSH, E. J., McDERMOTT, K. B., & ROEDIGER, H. L., III (2001). *Do retrieval factors play a role in the creation of false memories?* Manuscript submitted for publication.
- MATHER, M., HENKEL, L. A., & JOHNSON, M. J. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, **25**, 826-837.
- McDERMOTT, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language*, **35**, 212-230.
- McDERMOTT, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review*, **4**, 582-586.
- McDERMOTT, K. B., & ROEDIGER, H. L., III (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory & Language*, **39**, 508-520.
- McDERMOTT, K. B., & WATSON, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory & Language*, **45**, 160-176.
- McEVOY, C. L., NELSON, D. L., & KOMATSU, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1177-1194.
- MEYER, D. E., & SCHVANEVELDT, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, **90**, 227-234.
- MILLER, M. B., & WOLFORD, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, **106**, 398-405.
- NEELY, J. H. (1977). Semantic priming and retrieval from semantic memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, **106**, 226-254.
- NELSON, D. L., McEVOY, C. L., & SCHREIBER, T. A. (1999). *The University of South Florida word association, rhyme, and word fragment norms*. Unpublished manuscript, University of South Florida, Tampa.
- NELSON, D. L., SCHREIBER, T. A., & McEVOY, C. L. (1992). Processing implicit and explicit representations. *Psychological Review*, **99**, 322-348.
- 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.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, **76** (1, Pt. 2).
- 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.
- REYNA, V. F., & BRAINERD, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning & Individual Differences*, **7**, 1-75.
- ROEDIGER, H. L., III, BALOTA, D. A., & WATSON, J. M. (2001). Spreading activation and the arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95-115). Washington DC: American Psychological Association.
- 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. (1999). False alarms about false memories. *Psychological Review*, **106**, 406-410.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, **9**, 123-127.
- ROEDIGER, H. L., III, McDERMOTT, K. B., & PISONI, D. B. (2001). *Recollection of illusory voices*. Manuscript submitted for publication.
- 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. 187-245). Hove, U.K.: Psychological Press.
- 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). University of Minnesota, Minneapolis.
- SCHACTER, D. L., ISRAEL, L., & RACINE, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory & Language*, **40**, 1-24.
- 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.
- SEAMON, J. G., LUO, C. R., & GALLO, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science*, **9**, 20-26.
- SMITH, R. E., & HUNT, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, **5**, 710-715.
- STADLER, M. A., ROEDIGER, H. L., III, & McDERMOTT, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, **27**, 494-500.
- THAPAR, A., & McDERMOTT, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, **29**, 424-432.
- TOGLIA, M. P., & BATTIG, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.
- TOGLIA, M. P., NEUSCHATZ, J. S., & GOODWIN, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, **7**, 233-256.
- TULVING, E. (1974). Cue-dependent forgetting. *American Scientist*, **62**, 74-82.
- TULVING, E. (1979). Relation between encoding specificity and levels of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 405-428). Hillsdale, NJ: Erlbaum.
- TULVING, E. (1985). Memory and consciousness. *Canadian Psychologist*, **26**, 1-12.
- TUN, P. A., WINGFIELD, A., ROSEN, M. J., & BLANCHARD, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology & Aging*, **13**, 230-241.

- UNDERWOOD, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, **70**, 122-129.
- WELDON, M. S. (1991). Mechanisms underlying priming on perceptual tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 526-541.
- WICKENS, T. D., & HIRSHMAN, E. (2000). False memories and statistical decision theory: Comment on Miller and Wolford (1999) and Roediger and McDermott (1999). *Psychological Review*, **107**, 377-383.
- WIXTED, J. T., & STRETCH, V. (2000). The case against a criterion-shift account of false memory. *Psychological Review*, **107**, 368-376.
- ZECHMEISTER, E. B. (1969). Orthographic distinctiveness. *Journal of Verbal Learning & Verbal Behavior*, **8**, 754-761.

NOTES

1. The *orthographic distinctiveness* of a critical item was obtained by presenting 72 subjects with all 55 critical items visually and having them rate the words in accordance with procedures developed by Zechmeister (1969). Briefly, the subjects were instructed to rate each word for its physical appearance in terms of distinctiveness in the language on a 9-point scale, with 1 being *not very distinctive*, 5 corresponding to *average distinctiveness*, and 9 being *most distinctive*. To avoid ordering effects, three different randomized orderings of words were used across roughly equal groups of subjects. In the case of the critical item *sleep*,

for example, the distinctiveness is 5.15. Corresponding values for *butterfly* and for *man* are 7.44 and 2.13, respectively. Although orthographic distinctiveness was not included in our multiple regression analyses, this predictor is included in Appendix C.

2. The *orthographic neighborhood size*, or Coltheart's *N*, of each critical item was defined as the number of words that can be formed by changing a single letter in the word (Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, because it has no orthographic neighbors, Coltheart's *N* for the critical item *butterfly* is zero. In contrast, *sleep* has five orthographic neighbors (i.e., *sheep*, *sleek*, *sleet*, *steep*, and *sweep*). The reference set of words used were those in Kučera and Francis's (1967) norms. Although Coltheart's *N* was not included in our multiple regression analyses, this predictor is included in Appendix C.

3. *Word length*, *orthographic distinctiveness*, and *orthographic neighborhood size* of the critical item were all highly correlated. Specifically, the bivariate correlations among these three predictors were as follows: length and orthographic distinctiveness, $r = +.84$, $t(53) = 11.17$, $p < .05$; orthographic distinctiveness and orthographic neighborhood size, $r = -.66$, $t(53) = -6.40$, $p < .05$; length and orthographic neighborhood size, $r = -.63$, $t(53) = -5.82$, $p < .05$. Of these three variables, word length had the largest bivariate correlation with false recall ($r = -.37$), so it was the predictor used in the multiple regression analyses.

4. The multiple regression data matrix is available at www.iac.wustl.edu/nmcl/Web

APPENDIX A
The Fifty-Five 15-Word Lists and Their Associated Critical Items
With Backward Associative Strength (BAS) and Forward Associative Strength (FAS)
Values Used in the Multiple Regression Analysis of False Recall

Item	Word	BAS	FAS	Item	Word	BAS	FAS
Anger	mad	.393	.412	Black	white	.655	.557
	fear	.020	.059		dark	.111	.100
	hate	.028	.109		cat	.000	.043
	rage	.541	.042		charred	.023*	.000
	temper	.182	.000		night	.000	.021
	fury	.306	.000		funeral	.034	.000
	ire	.179*	.000		color	.074	.050
	wrath	.128*	.000		grief	.000	.000
	happy	.000	.042		blue	.028	.000
	fight	.034	.000		death	.016	.014
	hatred	.070*	.000		ink	.020	.000
	mean	.090	.000		bottom	.000	.000
	calm	.000	.000		coal	.288	.000
	emotion	.000	.000		brown	.338	.000
	enrage	.378	.000		gray	.365	.000
Mean	.157	.044	Mean	.130	.052		
Army	Navy	.543	.500	Bread	butter	.364	.487
	soldier	.287	.027		food	.000	.045
	United States	.000*	.000		eat	.000	.026
	rifle	.000	.000		sandwich	.067	.026
	Air Force	.133	.014		rye	.791	.000
	draft	.122	.000		jam	.054	.000
	military	.266	.027		milk	.012	.000
	Marines	.283	.047		flour	.142	.000
	march	.041	.000		jelly	.053	.019
	infantry	.284*	.000		dough	.310	.058
	captain	.014	.000		crust	.243	.000
	war	.000	.041		slice	.048	.019
	uniform	.048	.000		wine	.000	.000
	pilot	.000	.000		loaf	.552	.051
	combat	.000*	.000		toast	.364	.000
Mean	.135	.044	Mean	.200	.049		
Beautiful	ugly	.028	.229	Butterfly	moth	.267	.109
	pretty	.095	.389		insect	.000	.079
	girls	.033	.049		wing	.000	.030
	woman	.016	.014		bird	.000	.042
	homely	.000*	.000		fly	.000	.091
	lovely	.182*	.000		yellow	.000	.018
	nice	.000	.000		net	.000	.030
	picture	.000	.000		flower	.000	.036
	lady	.010	.000		bug	.000	.012
	mountain	.000	.000		cocoon	.412	.024
	snow	.000	.000		summer	.000	.000
	scene	.024	.000		color	.000	.030
	music	.000	.000		bee	.000	.000
	day	.000	.000		stomach	.000	.000
	gorgeous	.184	.056		worm	.000	.000
Mean	.038	.049	Mean	.045	.033		
Bitter	sweet	.020	.435	Cabbage	head	.000	.022
	sour	.115	.254		lettuce	.021	.281
	taste	.024	.065		vegetable	.000	.137
	chocolate	.000	.000		food	.000	.022
	rice	.000	.000		salad	.000	.022
	cold	.000	.101		green	.000	.079
	lemon	.000	.022		garden	.000	.000
	angry	.000	.000		leaf	.000	.029
	hard	.000	.000		sauerkraut	.042*	.000
	mad	.000	.000		smell	.000	.000
	acid	.000	.014		slaw	.041*	.043
	almonds	.000	.000		patch	.066	.115
	herbs	.000	.000		plant	.000	.000
	grape	.000	.000		carrots	.000	.000
	fruit	.000	.000		soup	.000	.014
Mean	.011	.059	Mean	.011	.051		

APPENDIX A (Continued)

Item	Word	BAS	FAS	Item	Word	BAS	FAS
Car	truck	.264	.111	City	town	.529	.307
	bus	.252	.022		crowded	.000*	.010
	train	.058	.011		state	.117	.132
	automobile	.709	.133		capital	.095	.000
	vehicle	.740	.000		streets	.054*	.046
	drive	.480	.122		subway	.000	.000
	jeep	.240	.000		country	.068	.020
	Ford	.331*	.000		New York	.383	.066
	race	.043	.011		village	.020	.000
	keys	.360	.000		metropolis	.536*	.000
	garage	.519	.000		big	.000	.025
	highway	.115	.000		Chicago	.152*	.000
	sedan	.510*	.000		suburb	.265	.010
	van	.448	.000		county	.195	.010
taxi	.129	.000	urban	.358*	.000		
	Mean	.347	.027		Mean	.185	.042
Carpet	rug	.468	.248	Cold	hot	.676	.413
	floor	.074	.159		snow	.199	.033
	soft	.000	.062		warm	.364	.033
	red	.000	.048		winter	.277	.022
	sweeper	.000*	.000		ice	.364	.098
	tack	.000	.000		wet	.108	.011
	walk	.000	.000		frigid	.570	.000
	bag	.000	.021		chilly	.395*	.000
	room	.000	.000		heat	.169	.000
	blue	.000	.021		weather	.032	.011
	chair	.000	.000		freeze	.461	.011
	thick	.000	.000		air	.000	.000
	deep	.000	.000		shiver	.669	.011
	magic	.014	.028		Arctic	.642	.000
wool	.000	.000	frost	.370	.000		
	Mean	.037	.039		Mean	.353	.043
Chair	table	.756	.314	Command	order	.000	.288
	sit	.183	.212		army	.000	.034
	legs	.000	.013		obey	.140	.062
	seat	.543	.109		officer	.000	.027
	couch	.288	.109		performance	.000	.014
	desk	.290	.019		do	.000	.014
	recliner	.547	.000		tell	.000	.055
	sofa	.132	.077		general	.000	.027
	wood	.012	.013		shout	.000	.000
	cushion	.086	.019		halt	.000	.000
	swivel	.593*	.000		voice	.000	.000
	stool	.320	.032		soldier	.000	.000
	sitting	.096	.000		harsh	.000	.000
	rocking	.593*	.019		attention	.000	.000
bench	.109	.013	sharp	.000	.000		
	Mean	.303	.063		Mean	.009	.035
Citizen	United States	.000*	.191	Cottage	house	.000	.381
	man	.000	.000		lake	.000	.063
	person	.000	.191		cheese	.000	.206
	American	.034	.086		home	.000	.040
	country	.000	.059		white	.000	.000
	alien	.000*	.079		cabin	.020	.016
	people	.000	.000		small	.000	.016
	vote	.000	.013		door	.000	.000
	me	.000	.000		fence	.000	.000
	patriot	.013	.020		vines	.000*	.000
	flag	.000	.000		woods	.000	.040
	foreigner	.000	.000		ivy	.000	.000
	France	.000	.000		roses	.000	.000
	immigrant	.000*	.013		cozy	.000	.000
member	.000	.033	hut	.030	.032		
	Mean	.003	.046		Mean	.003	.053

APPENDIX A (Continued)

Item	Word	BAS	FAS	Item	Word	BAS	FAS
Cup	mug	.268	.025	Fruit	apple	.154	.223
	saucer	.527	.418		vegetable	.220	.082
	tea	.054	.056		orange	.194	.174
	measuring	.385*	.000		kiwi	.709	.000
	coaster	.096*	.000		citrus	.426	.000
	lid	.000	.000		ripe	.151*	.000
	handle	.014	.000		pear	.347	.000
	coffee	.051	.105		banana	.215	.065
	straw	.029	.000		berry	.298	.000
	goblet	.118*	.000		cherry	.168	.000
	soup	.000	.062		basket	.084	.011
	stein	.014*	.000		juice	.035	.027
	drink	.011	.049		salad	.000	.000
	plastic	.075	.000		bowl	.028	.000
sip	.000	.000	cocktail	.000	.011		
	Mean	.109	.048		Mean	.202	.040
Doctor	nurse	.547	.379	Girl	boy	.701	.738
	sick	.031	.051		dolls	.199*	.000
	lawyer	.149	.101		female	.098	.013
	medicine	.152	.066		young	.000	.000
	health	.049	.020		dress	.063	.000
	hospital	.027	.015		pretty	.149	.027
	dentist	.214	.020		hair	.000	.000
	physician	.804	.040		niece	.026	.000
	ill	.000	.025		dance	.000	.000
	patient	.365	.025		beautiful	.049	.000
	office	.014	.010		cute	.035	.000
	stethoscope	.520	.000		date	.056	.000
	surgeon	.479	.040		aunt	.000	.000
	clinic	.300	.000		daughter	.042	.000
cure	.028	.010	sister	.041	.000		
	Mean	.245	.053		Mean	.097	.052
Flag	banner	.687	.000	Health	sickness	.220	.140
	American	.200	.269		good	.000	.133
	symbol	.014	.021		happiness	.000	.014
	stars	.000	.048		wealth	.022	.021
	anthem	.062*	.000		ill	.014	.000
	stripes	.177	.014		doctor	.020	.049
	pole	.157	.193		service	.000	.000
	wave	.000	.103		strong	.000	.000
	raised	.000*	.000		hospital	.000	.014
	national	.027	.000		disease	.000	.021
	checkered	.247*	.000		body	.027	.028
	emblem	.048*	.000		vigor	.000*	.000
	sign	.000	.000		center	.000	.000
	freedom	.021	.000		pain	.000	.000
pendant	.000*	.000	robust	.000*	.000		
	Mean	.109	.043		Mean	.020	.028
Foot	shoe	.321	.337	High	low	.777	.655
	hand	.158	.122		clouds	.000*	.000
	toe	.605	.235		up	.041	.034
	kick	.039	.000		tall	.000	.020
	sandals	.209	.000		tower	.079	.000
	soccer	.000	.000		jump	.072	.000
	yard	.126	.000		above	.057	.000
	walk	.016	.020		building	.000	.000
	ankle	.364	.000		noon	.033	.000
	arm	.000	.000		cliff	.028	.000
	boot	.142	.000		sky	.017	.014
	inch	.473	.020		over	.000	.000
	sock	.172	.000		airplane	.000	.014
	knee	.032	.000		dive	.028	.000
mouth	.000	.000	elevate	.174*	.000		
	Mean	.177	.049		Mean	.087	.049

APPENDIX A (Continued)

Item	Word	BAS	FAS	Item	Word	BAS	FAS
Justice	peace	.000	.151	Long	short	.222	.536
	law	.031	.171		fellow	.000	.000
	courts	.090	.158		narrow	.021	.000
	judge	.014	.096		John	.000*	.017
	right	.000	.021		time	.000	.034
	liberty	.113	.021		far	.041	.000
	government	.000	.000		hair	.103	.078
	jury	.000	.000		island	.000	.000
	truth	.082	.000		road	.020	.000
	blind	.000	.000		thin	.000	.011
	fair	.030	.027		underwear	.000	.000
	supreme	.021*	.000		distance	.150	.000
	crime	.014	.021		line	.031	.000
	department	.000	.000		low	.000	.000
	trial	.000	.000		rope	.000	.000
	Mean	.026	.044		Mean	.039	.045
	King	queen	.730		.772	Man	woman
England		.000	.000	husband	.018		.000
crown		.471	.016	uncle	.070		.000
prince		.134	.016	lady	.371		.013
George		.020	.000	mouse	.000		.000
dictator		.023	.000	male	.131		.000
palace		.159	.000	father	.048		.000
throne		.759	.000	strong	.020		.013
chess		.092	.000	friend	.000		.000
rule		.014	.031	beard	.055		.000
subjects		.000*	.000	person	.122		.000
monarch		.317	.039	handsome	.144		.000
royal		.315	.016	muscle	.048		.013
leader		.034	.000	suit	.074		.000
reign		.383*	.000	old	.034		.000
Mean		.230	.059	Mean	.115		.047
Lamp		light	.020	.769	Mountain		hill
	shade	.028	.058	valley		.195	.020
	table	.000	.019	climb		.291	.092
	bulb	.014	.045	summit		.108	.000
	post	.000	.026	top		.000	.041
	black	.000	.000	molehill		.256*	.031
	cord	.000	.000	peak		.248	.020
	desk	.034	.019	plain		.000	.000
	bright	.000	.000	glacier		.020	.000
	lighter	.000	.000	goat		.028	.000
	read	.000	.000	bike		.033	.000
	on	.000	.000	climber		.603	.031
	bed	.000	.000	range		.000	.051
	burn	.000	.013	steep		.061	.000
	stand	.000	.000	ski		.034	.000
	Mean	.006	.063	Mean		.154	.037
	Lion	tiger	.308	.362		Music	note
circus		.011	.000	sound	.205		.020
jungle		.034	.000	piano	.230		.020
tamer		.489*	.021	sing	.033		.088
den		.097*	.021	radio	.270		.041
cub		.063	.074	band	.432		.020
Africa		.014	.021	melody	.243		.020
mane		.200*	.021	horn	.014		.000
cage		.035	.000	concert	.395		.000
feline		.000*	.000	instrument	.148		.000
roar		.614	.032	symphony	.329		.000
fierce		.112	.021	jazz	.367		.000
bears		.034*	.021	orchestra	.309		.000
hunt		.000	.000	art	.020		.020
pride		.029	.000	rhythm	.277		.000
Mean		.136	.040	Mean	.227		.020

APPENDIX A (Continued)

Item	Word	BAS	FAS	Item	Word	BAS	FAS
Mutton	lamb	.024	.133	Rough	smooth	.416	.352
	sheep	.000	.027		bumpy	.150	.028
	meat	.000	.047		road	.000	.000
	chops	.000*	.000		tough	.192	.048
	beef	.000	.000		sandpaper	.429	.041
	veal	.000	.000		jagged	.128*	.000
	collar	.000	.000		ready	.000	.000
	leg	.000	.000		coarse	.291*	.014
	eat	.000	.000		uneven	.019	.000
	fat	.000	.000		riders	.027*	.000
	coat	.000	.000		rugged	.174*	.014
	stew	.000	.000		sand	.000	.000
	fur	.000	.000		boards	.000*	.000
	pork	.000	.000		ground	.000	.000
	steak	.000	.000		gravel	.000	.000
	Mean	.002	.014		Mean	.122	.033
Needle	thread	.758	.424	Rubber	elastic	.035*	.000
	pin	.289	.212		bounce	.018	.000
	eye	.000	.000		gloves	.033	.041
	sewing	.181*	.224		tire	.062	.095
	sharp	.030	.024		ball	.000	.041
	point	.024	.024		eraser	.026	.000
	prick	.108	.012		springy	.000*	.000
	thimble	.218	.000		foam	.116	.000
	haystack	.418	.030		galoshes	.063	.000
	thorn	.028	.000		soles	.000*	.041
	hurt	.000	.000		latex	.107*	.014
	injection	.331	.000		glue	.000	.000
	syringe	.520	.000		flexible	.041	.000
	cloth	.000	.000		resilient	.000*	.000
	knitting	.135	.000		stretch	.000	.027
	Mean	.203	.063		Mean	.033	.017
Pen	pencil	.476	.594	Shirt	blouse	.647	.135
	write	.128	.065		sleeves	.347*	.038
	fountain	.071	.000		pants	.185	.269
	leak	.000	.000		tie	.074	.103
	quill	.635	.000		button	.240	.064
	felt	.047	.000		shorts	.252	.013
	Bic	.372*	.000		iron	.010	.000
	scribble	.020	.000		polo	.177*	.000
	crayon	.000	.000		collar	.342	.032
	Cross	.013*	.000		vest	.143	.000
	tip	.000	.000		pocket	.058	.000
	marker	.257	.000		jersey	.174*	.000
	red	.000	.000		belt	.000	.000
	cap	.000	.000		linen	.000	.000
	letter	.000	.000		cuffs	.143*	.000
	Mean	.135	.044		Mean	.186	.044
River	water	.000	.071	Sleep	bed	.638	.092
	stream	.321	.118		rest	.475	.163
	lake	.142	.118		awake	.618	.143
	Mississippi	.654*	.031		tired	.493	.092
	boat	.000	.055		dream	.247	.194
	tide	.000	.000		wake	.304	.000
	swim	.000	.016		snooze	.520	.020
	flow	.283	.063		blanket	.024	.000
	run	.000	.016		doze	.682	.000
	barge	.047*	.000		slumber	.514	.000
	creek	.397	.000		snore	.439	.000
	brook	.161	.016		nap	.730	.000
	fish	.000	.016		peace	.000	.000
	bridge	.197	.000		yawn	.235	.000
	winding	.000	.016		drowsy	.551	.000
	Mean	.147	.036		Mean	.431	.047

APPENDIX A (Continued)

Item	Word	BAS	FAS	Item	Word	BAS	FAS
Slow	fast	.598	.527	Spider	web	.845	.246
	lethargic	.142*	.000		insect	.000	.127
	stop	.000	.034		bug	.040	.127
	listless	.000*	.000		fright	.000	.000
	snail	.486	.020		fly	.000	.016
	cautious	.027	.000		arachnid	.704*	.079
	delay	.059	.000		crawl	.000	.024
	traffic	.020	.000		tarantula	.744*	.000
	turtle	.372	.115		poison	.000	.000
	hesitant	.034*	.000		bite	.000	.000
	speed	.061	.014		creepy	.058*	.040
	quick	.272	.000		animal	.000	.000
	sluggish	.340*	.000		ugly	.000	.000
	wait	.000	.000		feelers	.000*	.000
	molasses	.170	.000		small	.000	.000
Mean	.172	.047	Mean	.159	.044		
Smell	nose	.108	.116	Stove	hot	.000	.285
	breathe	.000	.000		heat	.000	.030
	sniff	.442	.043		pipe	.000	.018
	aroma	.678	.000		cook	.000	.212
	hear	.000	.000		warm	.000	.012
	see	.000	.000		fire	.000	.000
	nostril	.000*	.000		oven	.224	.279
	whiff	.577*	.000		wood	.030	.000
	scent	.625	.029		kitchen	.056	.018
	reek	.510*	.000		lid	.000	.000
	stench	.562	.000		coal	.000	.000
	fragrance	.389	.000		gas	.026	.000
	perfume	.393	.036		iron	.000	.000
	salts	.028*	.000		range	.149	.012
	rose	.034	.000		furnace	.041*	.000
Mean	.290	.015	Mean	.035	.058		
Smoke	cigarette	.449	.323	Sweet	sour	.405	.372
	puff	.240	.000		candy	.336	.162
	blaze	.000	.000		sugar	.433	.061
	billows	.061*	.000		bitter	.435	.020
	pollution	.068	.000		good	.000	.014
	ashes	.052	.000		taste	.071	.014
	cigar	.507	.000		tooth	.000	.027
	chimney	.240	.000		nice	.095	.095
	fire	.018	.291		honey	.451	.000
	tobacco	.338	.000		soda	.000	.000
	stink	.000	.000		chocolate	.101	.041
	pipe	.419	.016		heart	.000	.000
	lungs	.119*	.000		cake	.027	.000
	flames	.000*	.000		tart	.223	.000
	stain	.000	.000		pie	.000	.000
Mean	.167	.042	Mean	.172	.054		
Soft	hard	.564	.509	Swift	fast	.016	.606
	light	.000	.012		slow	.000	.176
	pillow	.236	.018		river	.000	.000
	plush	.178	.000		Jonathan	.054*	.000
	loud	.333	.000		current	.000	.000
	cotton	.166	.018		rapid	.021	.012
	fur	.061	.000		stream	.000	.000
	touch	.061	.012		water	.000	.000
	fluffy	.266	.000		quick	.000	.048
	feather	.045	.024		Gulliver	.000*	.000
	furry	.061	.000		run	.000	.024
	downy	.221*	.000		sure	.000	.000
	kitten	.033	.000		deer	.000	.000
	skin	.161	.018		car	.000	.000
	tender	.297	.000		author	.000	.000
Mean	.179	.041	Mean	.006	.058		

APPENDIX A (Continued)

Item	Word	BAS	FAS	Item	Word	BAS	FAS
Thief	steal	.089	.388	Whistle	stop	.000	.032
	robber	.361*	.224		train	.045	.016
	crook	.459	.091		noise	.000	.071
	burglar	.257	.085		sing	.000	.016
	money	.000	.012		blow	.034	.175
	cop	.000	.000		tune	.000	.032
	bad	.000	.012		sound	.000	.000
	rob	.074	.000		dog	.000	.016
	jail	.013	.000		song	.000	.127
	gun	.000	.000		shrill	.000*	.000
	villain	.000	.000		boy	.000	.000
	crime	.028	.012		lips	.000	.024
	bank	.000	.000		wolf	.000	.000
	bandit	.167	.000		call	.000	.016
	criminal	.051	.012		loud	.000	.048
Mean	.100	.056	Mean	.005	.038		
Trash	garbage	.456	.526	Window	door	.156	.147
	waste	.067	.026		glass	.144	.256
	can	.014	.212		pane	.833	.179
	refuse	.017	.000		shade	.021	.058
	sewage	.053*	.000		ledge	.152	.013
	bag	.000	.026		sill	.682	.128
	junk	.126	.013		house	.000	.000
	rubbish	.397*	.013		open	.014	.019
	sweep	.000	.000		curtain	.189	.038
	scraps	.048*	.000		frame	.014	.013
	pile	.049	.000		view	.048	.026
	dump	.218	.013		breeze	.000	.000
	landfill	.186*	.000		sash	.000	.000
	debris	.266*	.000		screen	.027	.000
	litter	.209	.000		shutter	.480	.000
Mean	.140	.055	Mean	.184	.058		
Trouble	bad	.000	.123	Wish	want	.028	.071
	shooter	.000*	.048		dream	.014	.165
	worry	.000	.000		desire	.027	.039
	danger	.048	.027		hope	.068	.213
	sorrow	.000	.000		well	.000	.039
	fear	.000	.000		think	.000	.031
	school	.000	.000		star	.045	.126
	problem	.027	.089		bone	.000	.079
	police	.000	.041		ring	.000	.000
	fight	.000	.027		wash	.000	.016
	sad	.000	.000		thought	.000	.000
	difficulty	.031	.000		get	.000	.000
	help	.049	.041		true	.000	.016
	maker	.000	.000		for	.000	.000
	jail	.000	.000		money	.000	.000
Mean	.010	.026	Mean	.012	.053		
Whiskey	drink	.000	.081				
	drunk	.000	.121				
	beer	.000	.051				
	liquor	.000	.081				
	gin	.021	.000				
	bottles	.000	.020				
	alcohol	.000	.051				
	rye	.000	.000				
	glass	.000	.020				
	wine	.000	.030				
	rum	.035	.081				
	bourbon	.144	.051				
	evil	.000	.000				
	bar	.000	.000				
	scotch	.135	.040				
Mean	.022	.042					

Note—Unless otherwise indicated with an asterisk, all BAS and FAS values are drawn from the associative norms of Nelson, McEvoy, and Schreiber (1999). Those values shown with an asterisk were obtained by the authors by using norming procedures similar to those of Nelson et al. (1999). The corresponding mean for these BAS and FAS values for each list are also represented.

APPENDIX B
Data Matrix of the Fifty-Five 15-Word Lists and Associated Critical Items
Used in the Multiple Regression Analysis of False Recall

Critical Item	False Recall	False Recgn	Length	Log Freq	Concrete	FAS	BAS	Connect	Vrd Recall	Raw Freq	Ortho Dist	Colt <i>N</i>
anger	.490	.790	5	1.69	3.75	.044	.157	1.73	.500	48	4.30	3
army	.250	.530	4	2.12	6.53	.044	.135	2.53	.610	132	4.12	2
beautiful	.030	.440	9	2.11	3.89	.049	.038	2.13	.710	127	6.38	0
bitter	.010	.260	6	1.73	4.05	.059	.011	1.40	.680	53	5.78	10
black	.340	.490	5	2.31	4.66	.052	.130	1.47	.600	203	4.41	3
bread	.310	.640	5	1.62	6.18	.049	.200	1.33	.550	41	3.97	5
butterfly	.010	.260	9	0.40	5.91	.033	.045	1.93	.700	2	7.44	0
cabbage	.050	.440	7	0.65	6.07	.051	.011	2.53	.650	4	6.99	0
car	.350	.420	3	2.44	6.35	.027	.347	1.73	.660	274	2.08	19
carpet	.150	.490	6	1.13	5.68	.039	.037	0.60	.610	13	4.37	0
chair	.540	.740	5	1.82	6.12	.063	.303	1.93	.640	66	3.60	3
citizen	.100	.600	7	1.48	4.51	.046	.003	2.27	.680	30	6.26	0
city	.460	.640	4	2.59	5.41	.042	.185	1.80	.650	393	4.38	3
cold	.440	.840	4	2.23	4.67	.043	.353	2.27	.610	171	3.60	14
command	.080	.550	7	1.86	n.a.	.035	.009	0.67	.580	72	5.82	1
cottage	.080	.380	7	1.29	5.93	.053	.003	1.00	.660	19	6.52	0
cup	.450	.820	3	1.66	5.35	.048	.109	1.53	.530	45	2.66	7
doctor	.600	.710	6	2.00	5.75	.053	.245	2.60	.570	100	5.00	0
flag	.310	.600	4	1.22	6.20	.043	.109	1.00	.630	16	4.25	5
foot	.350	.620	4	1.85	3.46	.049	.177	1.40	.640	70	5.18	10
fruit	.200	.450	5	1.55	6.00	.040	.202	1.67	.710	35	4.58	0
girl	.320	.580	4	2.34	6.83	.052	.097	1.27	.670	220	4.15	2
health	.110	.550	6	2.02	3.54	.028	.020	1.27	.620	105	4.97	2
high	.260	.720	4	2.70	3.62	.049	.087	1.47	.580	497	5.18	3
justice	.300	.760	7	2.06	2.18	.044	.026	1.40	.590	114	5.67	1
king	.100	.270	4	1.95	5.54	.059	.230	2.07	.650	88	4.40	9
lamp	.140	.630	4	1.27	6.09	.063	.006	0.73	.610	18	3.51	9
lion	.230	.330	4	1.24	6.14	.040	.136	0.80	.630	17	4.00	6
long	.030	.340	4	2.88	3.68	.045	.039	0.47	.600	755	3.36	10
man	.240	.610	3	3.08	6.14	.047	.115	0.80	.560	1207	2.13	33
mountain	.420	.690	8	1.53	6.25	.037	.154	0.87	.600	33	5.62	1
music	.340	.690	5	2.34	5.15	.020	.227	1.60	.590	216	3.71	0
mutton	.010	.110	6	0.93	5.34	.014	.002	2.00	.650	8	6.37	4
needle	.520	.680	6	1.19	5.79	.063	.203	1.93	.600	15	6.29	0
pen	.350	.570	3	1.27	5.59	.044	.135	1.33	.630	18	2.15	19
river	.420	.670	5	2.22	5.83	.036	.147	2.47	.640	165	3.68	6
rough	.530	.830	5	1.62	4.48	.033	.122	1.27	.560	41	4.74	8
rubber	.320	.670	6	1.19	6.04	.017	.033	0.67	.530	15	5.89	2
shirt	.270	.540	5	1.44	6.05	.044	.186	1.40	.640	27	3.84	4
sleep	.610	.800	5	1.82	4.74	.047	.431	1.80	.610	65	5.15	5
slow	.420	.690	4	1.78	2.89	.047	.172	1.13	.530	60	3.38	10
smell	.600	.840	5	1.54	4.40	.015	.290	1.47	.580	34	4.32	5
smoke	.540	.730	5	1.62	5.16	.042	.167	1.80	.640	41	5.05	2
soft	.460	.810	4	1.79	4.10	.041	.179	0.80	.590	61	3.70	4
spider	.370	.580	6	0.40	5.95	.044	.159	1.33	.620	2	5.78	0
stove	.180	.700	5	1.19	5.75	.058	.035	2.00	.600	15	4.29	7
sweet	.540	.780	5	1.85	4.53	.054	.172	2.07	.630	70	5.47	6
swift	.140	.350	5	1.51	3.31	.058	.006	1.73	.630	32	5.43	1
thief	.230	.700	5	0.93	4.83	.056	.100	4.27	.610	8	5.34	1
trash	.490	.780	5	0.40	5.76	.055	.140	2.60	.540	2	4.00	2
trouble	.080	.540	7	2.13	2.25	.026	.010	1.07	.580	134	5.53	1
whiskey	.030	.530	7	1.24	6.00	.042	.022	4.93	.720	17	7.70	1
whistle	.090	.190	7	0.65	5.58	.038	.005	1.13	.600	4	6.52	0
window	.650	.840	6	2.08	6.27	.058	.184	0.67	.630	119	5.73	1
wish	.290	.800	4	2.04	2.66	.053	.012	0.87	.580	110	4.15	6

APPENDIX C
Connectivity Matrix for the Critical Item *Sleep*

Cue	Target															Sum
	bed	rest	awake	tired	dream	wake	snooze	blanket	doze	slumber	snore	nap	peace	yawn	drowsy	
bed	ND	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
rest	1	ND	0	1	0	0	0	0	0	0	0	0	1	0	0	3
awake	0	0	ND	1	0	0	0	0	0	0	0	0	0	0	0	1
tired	1	0	1	ND	0	0	0	0	0	0	0	0	0	0	1	3
dream	0	0	0	0	ND	0	0	0	0	0	0	0	0	0	0	0
wake	0	0	1	0	0	ND	0	0	0	0	0	0	0	0	0	1
snooze	0	0	0	1	0	0	ND	0	1	0	1	1	0	0	0	4
blanket	1	0	0	0	0	0	0	ND	0	0	0	0	0	0	0	1
doze	0	0	0	1	0	0	1	0	ND	0	1	1	0	0	0	4
slumber	0	0	0	1	0	0	0	0	0	ND	0	1	0	0	0	2
snore	0	0	1	0	0	0	0	0	0	0	ND	0	0	0	0	1
nap	1	1	0	1	0	0	0	0	0	0	0	ND	0	0	0	3
peace	0	0	0	0	0	0	0	0	0	0	0	0	ND	0	0	0
yawn	0	0	0	1	0	0	0	0	0	0	0	0	0	ND	0	1
drowsy	1	0	0	1	0	0	0	0	0	0	0	0	0	0	ND	2
Sum	5	1	3	9	0	0	1	0	1	0	2	3	1	0	1	$N = 27$

Mean = 1.8

(Manuscript received July 10, 2000;
revision accepted for publication February 23, 2001.)