

Toward a model of false recall: Experimental manipulation of encoding context and the collection of verbal reports

KERRI A. GOODWIN

Kalamazoo College, Kalamazoo, Michigan

and

CHRISTIAN A. MEISSNER and K. ANDERS ERICSSON

Florida State University, Tallahassee, Florida

The likelihood of false recall in the Deese–Roediger–McDermott (DRM) paradigm was shown to depend on encoding context in two experiments. When fillers had been preselected to decrease the likelihood of encoding the critical lure’s semantic features, false recall was virtually eliminated. However, when the same words were presented rearranged in different presentation orders, levels of false recall that were found in earlier DRM studies (Robinson & Roediger, 1997) were replicated. The role of encoding processes in the DRM paradigm was further explored with additional participants completing the experiment while thinking aloud. During encoding of word lists, participants verbalized semantic elaboration of the critical lure while studying the word lists. A path analysis demonstrated that participants’ verbalization of critical lures during encoding reliably predicted their level of false recall.

It is well known that human memory is fallible; however, Roediger and McDermott (1995) demonstrated laboratory conditions that evoke high levels of false recall in a list-learning paradigm. For example, when participants study lists of words, such as *thread*, *pin*, *eye*, *sewing*, and so on, during subsequent recall they also tend to remember the associated critical lure *needle*, although this word had never been presented to them. This general paradigm was first developed by Deese (1959) and is currently referred to as the *Deese–Roediger–McDermott* (DRM) paradigm. In a recent extensive review, Roediger, McDermott, and Robinson (1998) stated that while previous researchers have noted the robust and reliable nature of this false memory effect, including the large number of variables that affect false recall, they acknowledge that the processes mediating the effect are still poorly understood. The goal of the present set of experiments was to increase our knowledge of the cognitive processes that mediate encoding and retrieval of words that have never been presented. In this

study, we explored two different approaches. First, we proposed new experimental manipulations that reduce the likelihood of false recall. Second, we examined the cognitive processes associated with false recall by collecting concurrent verbal reports during encoding and recall of list items. Before turning to a description of our experiments, we will first review the body of relevant findings that will constrain any model of the DRM effect.

Phenomenological Aspects of False Recall

When questioned after the completion of recall in the DRM paradigm, participants report high levels of confidence in their memory for having “heard” the nonpresented items (Roediger & McDermott, 1995, Experiment 1) and recount various details regarding the occurrence of the nonpresented lure during encoding (Roediger & McDermott, 1995, Experiment 2). Norman and Schacter (1997) have also found that individuals reported remembering semantic information about the nonpresented items. For example, in retrospective reports of nonpresented lures recognized as “old,” participants reported the following: “needle: this word came in the same list as thread”; “music: I thought about music I heard this morning”; and “thief: heard word ‘stop,’ thought of screaming ‘stop thief’” (quoted in Norman & Schacter, 1997, p. 841).

McDermott (1997) argued that critical lures consciously come to mind during the original encoding of list items, because she found evidence for perceptual priming of critical lures during study of the DRM word lists. Her interpretation is supported by Roediger and McDermott’s

The authors express their gratitude to Charles Brainerd, Kathleen McDermott, and Mary Susan Weldon for their useful comments, suggestions, and recommendations on earlier versions of this article. Thanks also to Richard Wagner for his statistical consultations. The authors also thank Colleen Kelley and Rolf Zwaan for their constructive comments early in the course of this project. Special thanks to Roxana Gonzalez, Melissa Christle, and Ryan Heintz for their assistance in the data collection phase of these experiments and in the transcriptions of verbal reports. Correspondence should be addressed to K. A. Goodwin, Department of Psychology, Kalamazoo College, 1200 Academy St., Kalamazoo, MI 49006 (e-mail: kgoodwin@kzoo.edu).

Table 1
Example of *soft* List in Eights, Fours, Twos, Ones, and Contextually Unrelated (CU) List Structures

8-Related, 8-Unrelated	4-Related, 4-Unrelated	2-Related, 2-Unrelated	1-Related, 1-Unrelated	CU
hard	hard	hard	hard	hard
light	light	light	<i>tone</i>	<i>hat</i>
pillow	pillow	<i>tone</i>	light	light
plush	plush	<i>trade</i>	<i>trade</i>	<i>bulb</i>
loud	<i>tone</i>	pillow	pillow	pillow
cotton	<i>trade</i>	plush	<i>gin</i>	<i>case</i>
fur	<i>gin</i>	<i>gin</i>	plush	plush
touch	<i>speaker</i>	<i>speaker</i>	<i>speaker</i>	<i>garden</i>
<i>hat</i>	<i>garden</i>	loud	loud	loud
<i>bulb</i>	<i>case</i>	cotton	<i>garden</i>	<i>speaker</i>
<i>case</i>	<i>bulb</i>	<i>garden</i>	cotton	cotton
<i>garden</i>	<i>hat</i>	<i>case</i>	<i>case</i>	<i>gin</i>
<i>speaker</i>	loud	fur	fur	fur
<i>gin</i>	cotton	touch	<i>bulb</i>	<i>trade</i>
<i>trade</i>	fur	<i>bulb</i>	touch	touch
<i>tone</i>	touch	<i>hat</i>	<i>hat</i>	<i>tone</i>

Note—Unrelated filler items are given in italics.

(1995) earlier findings of high levels of “remember” judgments in the false recognition of critical lures (see also Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996). Taken together, these phenomenological findings suggest that memory associated with the DRM effect is based on memory traces generated prior to the recall phase.

Encoding Manipulations in the DRM Paradigm

Other researchers have successfully influenced the likelihood of false recall and recognition by manipulating the conditions of encoding. For example, Robinson and Roediger (1997) found that the probability of false recall was increased when the number of list items associated with the critical lure was increased. Read (1996) instructed participants to think about and rehearse one list of words that would help them in a later test of word meanings (referred to as *elaborative rehearsal*, p. 107). These semantic encoding instructions led to increases in false recall of the critical lure *sleep* relative to instructions to encode list items in their serial order. Furthermore, McDermott (1996) and Toglia, Neuschatz, and Goodwin (1999) found that by presenting all list words associated with the critical lure together in blocked groups produced a higher probability of false recall than when all words were randomly distributed across the list. Finally, in an experiment in which depth of processing was manipulated, Toglia et al. found a higher likelihood of false recall when participants encoded semantic properties of the presented words than when they encoded surface properties of the presented words. Thus, the effect of an experimental manipulation of encoding context on false recall suggests that the processes at encoding contribute to access and later false recall of critical lures.

Toward a Model of False Recall

Our analysis of the research literature suggests that the phenomenon of false recall involves memory traces generated during the encoding of list words. Consistent with participants' phenomenological reports of the semantic

attributes of falsely recognized items (Norman & Schacter, 1997), we propose that critical lures are explicitly accessed during the encoding of items in the list when the words are semantically processed (Toglia et al., 1999). Furthermore, the probability of their retrieval increases when several associates of the lure are presented together, a procedure that appears to encourage the semantic grouping and elaboration of list items (cf. the blocked condition in McDermott, 1996). Thus, during recall, participants may be unable to distinguish between semantic elaborations associated with presented words and those associated with the critical lure. Consequently, participants are more likely to falsely recall the critical lure as a result of previously accessing it during encoding.

Robinson and Roediger (1997) found that increasing the number of list associates in the DRM lists led to higher levels of false recall of critical lures. We hypothesized that the contextual organization of list structure could be changed without altering the number of list items in order to influence the likelihood that the critical lure is accessed during encoding and falsely recalled. To test these predictions, we designed lists in which the encoding context of presented words was systematically varied. Similar to Robinson and Roediger, we constructed lists of eight associates and eight fillers. However, our filler items were specifically chosen to provide contextual associations that were unrelated to the overall list theme or critical lure. For example, in the *soft* list (in which *soft* was the critical lure), the associate *hard* was followed by the filler *hat*, the associate *pillow* was followed by the filler *case*, and the associate *loud* was followed by the filler *speaker*. Subsequently, five list structures containing the same words in differing orders were created as a manipulation of contextual organization of the lists: (1) *eights*, 8 associates and 8 fillers; (2) *fours*, 4 associates, 8 fillers, and 4 associates; (3) *twos*, 2 associates and 2 fillers, alternating; (4) *ones*, 1 associate and 1 filler unrelated to previous associate, alternating; and (5) the list structure *contextually unrelated* (CU) to the lure (see Table 1).

In the eights, fours, twos, and ones list structures, the filler items were positioned such that they were not in proximity to their related associate. Thus, semantic relations between the list associates were likely to be evoked, including generation of the critical lure. In the CU list structure, each associate was immediately followed by (or paired with) its related filler item, which was predicted to minimize activation of the overall list theme and critical lure during encoding.

We proposed that the presentation of fillers that guide semantic encoding away from the critical lure (CU list structure) would reduce the frequency of explicit access to the critical lure and subsequently reduce the likelihood of false recall. McDermott (1996) and Toglia et al. (1999) found that randomly ordered list items reduced false recall relative to lists that were blocked by theme. In our experiments, all lists were blocked by theme, and only in the CU list structure condition were items contextually manipulated in an effort to reduce false recall in the face of the list theme. The other list conditions assessed the effect of physical proximity of the list words associated with the critical lure. Here, we hypothesized that the encoding of the critical lure would more likely occur in lists that were organized around the critical theme, specifically when themed items were in close proximity (cf. Robinson & Roediger, 1997). Therefore, themed items in close proximity to each other (as in the eights list structure) would lead to a higher likelihood of false recall. In a companion experiment, we collected "think aloud" protocols from a small number of participants in an effort to trace their encoding and retrieval processes. We hypothesized that false recall in the standard DRM conditions is related to verbalized encodings involving the lure generated during the presentation of the lists.

In sum, we have proposed a two-pronged approach to the study of false recall using the DRM paradigm. First, in two experiments, we experimentally manipulated the encoding context of list items in an attempt to eliminate false recall. Second, using the procedures from these experiments, we further explored the encoding and retrieval processes involved in the generation and false recall of critical lures with the collection of verbal reports. Furthermore, we explored path models of false recall, in which verbalized encoding processes predict the likelihood of false recall.

EXPERIMENT 1

The purpose of Experiment 1 was to investigate the likelihood of false recall when contextual organization of list structure was systematically manipulated. To reiterate, if false recall is due to individuals' elaborative encoding of list words and critical lures, as indicated from levels of processing approaches to the DRM effect (Toglia et al., 1999), then the effects of experimentally induced contextual organization of list words should differentially influence the probability of false recall. Specifically, it

was predicted that the CU list structure would lead to a lower rate of false recall than would all other list structures, since this list structure was designed to induce an encoding context unrelated to the critical lure. Additionally, it was predicted that the CU list structure would lead to the highest level of correct recall relative to all other list structures, because the structural organization of the CU list should lend itself to more distinctive encodings of the list items as opposed to more elaborative, error-prone encodings of a list theme. Furthermore, as words associated to the lure become distributed across list structures (eights, fours, twos, ones), it was hypothesized that probabilities of true and false recall should decrease.

Method

Participants. A total of 64 undergraduates participated in the experiment to meet course requirements for an introductory psychology course. The participants were tested in a computer lab at computer terminals. They were tested in groups ranging in size from 10 to 20 participants.

Materials. Ten 16-item word lists were created for use in the experiment, each list containing the first 8 highly associated words drawn from the original lists prepared by Roediger and McDermott (1995). Selected lists included *bread, chair, doctor, foot, mountain, needle, rough, sleep, soft, and sweet*. Eight filler words were chosen for each list by the experimenters in the formation of the CU list structure. Fillers were chosen with the aid of an Oxford Wordfinder (Reader's Digest Editorial Staff, 1996) due to their association or relation with a selected word from the original 8 associates, but not to the overall theme or critical lure. For example, the associate *hard* was paired with the filler *hat*, which has little or no relation to the critical lure *soft*. Following the selection of all fillers, lists were ordered according to the five structures outlined earlier, each maintaining the same sets of words across the manipulation (see Table 1).

Four random orders for the presentation of list structure and four random orders of list themes (e.g., *sleep, chair, soft*) were created, after which the total presentation was counterbalanced to control for order effects.¹ Two additional lists were prepared as practice lists to familiarize the participants with the experimental procedures. Unlike the 10 target lists, the training lists contained words unrelated to each other or to an overall list theme.

Design and Procedure. The present experiment was designed as a replication and extension of Robinson and Roediger's (1997) study and used the same procedures whenever possible. The participants were presented with 12 lists of 16 words each. List structure was manipulated as a within-subjects variable; the participants were presented two lists from each of the five list structures resulting in 10 experimental lists. Two practice lists were first presented to introduce the participants to the memory task and the computer-style presentation and were not included in the data analysis.

Instructions appeared on a computer screen and were read aloud by the experimenter. The participants were instructed that the experiment involved memory for lists of words and that they should pay close attention to the presented words for a subsequent memory test. All word lists and instructions were presented on Power Macintosh 7200/120 personal computers. Programs were timed to present each word for 2 sec, followed by a blank screen for 2 sec until the next word was presented. During a 30-sec delay from the initial presentation of each list, several three-digit multiplication problems were displayed for the participants to solve. After the multiplication problems, the participants were instructed to recall the list words and were instructed not to guess. All participants attempted to recall each list for 1.5 min. The participants recorded

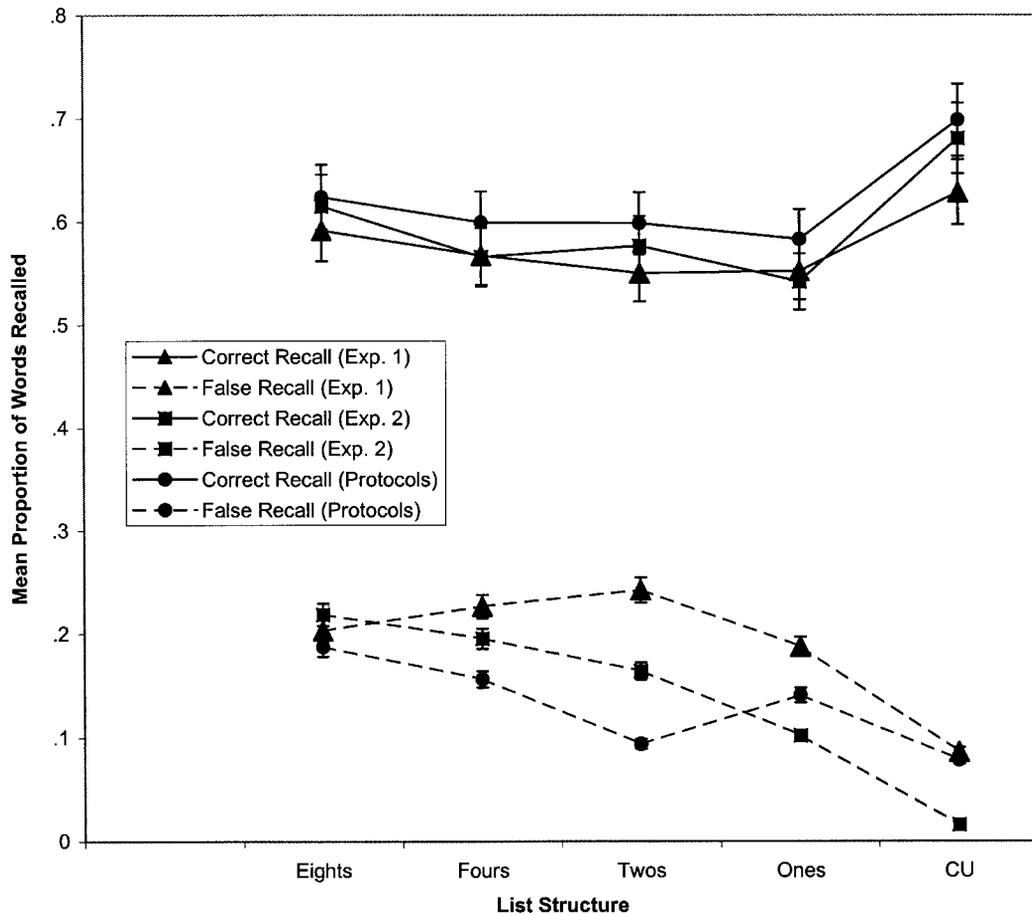


Figure 1. Mean proportions of true and false recall as a function of list structure for the sequential presentation of list items for Experiments 1 and 2 and the verbal protocol sample.

their recall responses in booklets that were provided by the experimenters. This procedure was repeated for all 12 lists.

Results

Within-subjects analyses of variance (ANOVAs), with list structure as a repeated measure, were used to analyze both true and false recall performance. Figure 1 displays the proportion of true and false recall as a function of list structure.

False recall. List structure significantly influenced the likelihood of false recall of critical lures [$F(4,252) = 3.42$, $MS_e = 0.07$, $p < .05$]. Consistent with our predictions, the CU list structure led to the lowest levels of false recall relative to all other list structures, as illustrated in Figure 1. Planned pairwise comparisons confirmed that false recall for the CU list structure ($M = .09$, $SD = .19$) was reliably lower than false recall in all other list structures, including the ones condition ($M = .19$, $SD = .27$), which closely matched the structure of the CU list (all $ps < .05$). Additionally, all list structures, except the CU list, produced false recall rates comparable to Robinson and Roediger's (1997) 9-associates, 6-fillers condition ($M = .19$). A separate within-subjects ANOVA, exclud-

ing the CU list structure, was conducted to determine whether false recall decreased when list associates were distributed across the list. However, the analysis did not produce a significant effect of the distribution of list associates across list structure as originally predicted, and, therefore, no linear trend across list structures was observed.

Correct recall. The main effect of list structure was reliable for correct recall of list words (see Figure 1) [$F(4,252) = 8.83$, $MS_e = 0.008$, $p < .001$]. As predicted, pairwise comparisons demonstrated that the proportion of correct recall was highest for the CU list structure ($M = .63$, $SD = .15$) relative to all other list structures (all $ps < .05$). When the data for the CU list structure were excluded, another within-subjects ANOVA still revealed a significant main effect for list structure [$F(3,189) = 3.12$, $MS_e = 0.007$, $p < .05$]. To test our hypothesis that correct recall would decrease as the words associated to the lure were distributed across list structures, we conducted a trend analysis with linear, quadratic, and cubic trends. Only the linear trend for eights, fours, twos, and ones list structures was reliably related to correct recall [$F(1,189) = 7.11$, $MS_e = 0.008$, $p < .05$].

Summary. The proportions of both true and false recall were significantly influenced by manipulations of the contextual organization of list structure, particularly with regard to the CU list structure. All list structures (except the CU list structure) led to representative levels of false recall with means ranging from .19 to .24, comparable to Robinson and Roediger's findings for 9 items and 6 fillers ($M = .19$). The CU list structure that contained filler items selected to bias the semantic encoding of list items away from the critical lure produced reliably lower levels of false recall than all other list types, including the ones list structure, which had similar spacing of list associates. The elevated level of correct recall for the CU list structure demonstrates that the reduced level of false recall is not an artifact of poor encoding of list items in memory but suggests a more qualitative change in the semantic encoding of the list items. The theoretical implications of a dissociation between true and false recall in the CU list structure will be discussed later in the context of theoretical explanations for false recall.

EXPERIMENT 2

The results of Experiment 1 demonstrated that it is possible to manipulate the encoding context of words and dramatically reduce the frequency of false recall. As part of our second approach to study the encoding and retrieval processes of the DRM effect, we collected additional think-aloud protocols in Experiment 1 (to be discussed in more detail later). These protocols showed that the participants often elaborated on list items as soon as the words appeared on the computer screen and began to encode associations to the presented words before the next item appeared on the screen. As a result, some participants verbalized elaborate associations to the critical lure before the biasing filler item in the CU condition appeared on the screen and had the opportunity to influence the memory encoding of words associated to the lure. To further control contextual encoding of list items in the CU list structure, we presented list associates (e.g., *hard*) together with the biasing filler item (e.g., *hat*) in a paired presentation format. All other list structures were also presented in a paired presentation format. Again, we predicted that the CU list structure would lead to the lowest level of false recall and the highest level of correct recall relative to all other list structure conditions. Furthermore, we predicted that, as list items become distributed across the eights, fours, twos, and ones list structures, false recall and correct recall should decrease. In an attempt to trace encoding and retrieval processes in the DRM paradigm as in Experiment 1, additional participants were asked to "think aloud" using the procedure in Experiment 1. These results will be discussed in a separate section.

Method

Participants. Sixty-four undergraduates participated in the experiment to meet course requirements for an introductory psychology course. They were tested in a computer lab at computer terminals in groups ranging in size from 10 to 20 participants.

Materials, Design, and Procedure. The word lists and randomization schedule used were the same as in Experiment 1, with the exception that list items were paired rather than presented sequentially. Word pairs were presented for 4 sec, followed by a 4-sec pause before another word pair was shown. This timing was equivalent to the total presentation time for the same two words presented sequentially in Experiment 1. Again, list structure was manipulated as a within-subjects factor, in which each participant was exposed to each of the five list structures on two occasions. All other aspects of the procedure remained the same as in Experiment 1.

Results

Within-subjects ANOVAs, with list structure as a repeated measure, were used to analyze both true and false recall performance.

False recall. As shown in Figure 1, results demonstrated that list structure significantly influenced the likelihood of false recall of critical lures [$F(4,252) = 8.32$, $MS_e = 0.05$, $p < .001$]. Consistent with our original hypotheses, pairwise comparisons indicated that the likelihood of false recall was significantly reduced, if not virtually eliminated, in the CU list structure ($M = .02$, $SD = .09$) relative to all other list structures (all $ps < .01$). These findings support the original hypothesis that the CU list structure should lead to a decrease in false recall of critical lures. Consistent with Experiment 1, all list structures except the CU list structure produced false recall rates comparable to those observed by Robinson and Roediger (1997) for their 9-associates, 6-fillers condition ($M = .19$).

To test our prediction that false recall would decrease as list associates became distributed across list structures, we conducted a separate within-subjects ANOVA with the CU list structure excluded. There was a significant main effect for list structure [$F(3,189) = 2.70$, $MS_e = 0.06$, $p < .05$]. A trend analysis for linear, quadratic, and cubic trends showed that only the linear trend for the eights, fours, twos, and ones conditions was significant [$F(1,189) = 7.44$, $MS_e = 0.06$, $p < .01$]. Hence, false recall decreased as list associates were distributed across list structure conditions.

Correct recall. As shown in Figure 1, results also indicated a significant main effect for list structure for correct recall of list words [$F(4,252) = 26.05$, $MS_e = 0.007$, $p < .001$]. As predicted, pairwise comparisons demonstrated that the proportion of correct recall was highest for the CU list structure ($M = .68$, $SD = .13$), which was reliably higher than all other list structures (all $ps < .05$) including the eights condition ($M = .62$, $SD = .13$; $p < .05$). When the data from the CU list structure were excluded, a separate within-subjects ANOVA revealed a significant main effect for list structure [$F(3,189) = 9.01$, $MS_e = 0.007$, $p < .001$]. As predicted, there was a reliable linear trend, which demonstrated that, as list items became distributed across eights, fours, twos, and ones list structures, correct recall decreased [$F(1,189) = 18.73$, $MS_e = 0.007$, $p < .001$]. However, this was further qualified by a significant cubic trend [$F(1,189) = 8.14$, $MS_e = 0.004$, $p < .01$].

Summary. As in Experiment 1, the proportions of both true and false recall were significantly influenced by ma-

nipulations to the contextual organization of list items, particularly with regard to the CU list structure. Furthermore, with a paired presentation of list items, there were decreases in both true and false recall with the distribution of list items across list structures. Again, a dissociation between true and false recall was demonstrated in the CU list structure. Namely, for the CU list structure, correct recall exceeded that of the eight list structure, yet false recall was virtually eliminated. Again, these data suggest a qualitative change in the semantic encoding of list items, particularly in the CU list structure condition, which will be discussed later.

VERBAL PROTOCOL EXPERIMENT

In Experiments 1 and 2, we found that experimental manipulations designed to decrease explicit access to the critical lure during the encoding of list items reduced the level of false recall. In our second approach to the study of false recall in the DRM paradigm, we assessed the cognitive processes associated with false recall by instructing different samples of participants to “think aloud” during both the study and recall of the presented lists using the experimental procedures from either Experiment 1 or Experiment 2. If the participants explicitly generated the critical lure during study by producing encodings of list items based on associations to it, we would expect that participants would be more likely to verbalize the lexical term or some semantic referent as part of their concurrent verbalizations. The think-aloud participants studied list items with the same presentation rates and formats as in Experiments 1 and 2, with the additional instruction to verbalize their spontaneous thoughts during the encoding and recall of list items. Given that the think-aloud participants were instructed to verbalize their thoughts in real time, one cannot assume that all thoughts are verbalized. The potential incompleteness of verbal reports limits theoretical claims to the structure and frequency of the verbalized thoughts. Thus, we only make inferences of encoding processes based on the thoughts that the participants verbalized during encoding.

We proposed that the think-aloud protocols obtained during the participants’ study would reflect some of the encoding and retrieval processes that eventually result in false recall of critical, nonpresented lures. Furthermore, assessment of on-line verbal protocols during encoding of word lists in the DRM paradigm would provide rich data for predicting false recall. Roediger et al. (1998) noted that the DRM effect may occur because “subjects are consciously, effortfully processing and rehearsing these non-presented items during the study phase” (p. 212). Therefore, the present investigation assessed three aspects of the think-aloud protocols during encoding: (1) type of encoding strategy, (2) whether or not the critical lure was verbalized during encoding, and (3) the number of unique contexts in which the critical lure was verbalized. At recall, think-aloud protocols were also coded for (4) whether or not the critical lure was verbalized during recall and

(5) whether or not the critical lure was falsely recalled. Appendix A provides a brief summary of each coding variable. Two of the more subjective coding categories (encoding strategy and contextual variability) are discussed in more detail below.

Two major types of verbal reports are expected to occur during the study of list items in any memory experiment, including the DRM paradigm: semantic elaboration and verbatim rehearsal. For example, when participants think about the presented words for the purpose of memorizing them, they may create a complex context that links the concepts in a meaningful manner (semantic elaboration), or they may merely repeat the words to themselves without any additional semantic information (verbatim rehearsal). Unfortunately, it is possible that participants encode words semantically yet still only verbalize the words verbatim. On the other hand, if the participants merely rehearsed the items verbatim, they would not verbally report appropriate contexts for the words. Following the practice of coding verbal reports, we distinguished between two encoding strategies used by the participants. We defined the category of *semantic elaboration* as verbalizations involving the on-line elaboration of list items in which information beyond or other than the list items themselves was present (e.g., “putting *thread* on the *sharp pin*”). Although the verbalization of the critical lure during encoding may constitute a semantic elaboration in and of itself, we examined the overall encoding strategy used by the participants. We defined the category of *verbatim rehearsal* as verbalizations of list words without information other than the list items themselves (e.g., “*thread, pin, sharp*”). Additionally, each time the critical lure is verbalized during encoding, it is possible that it may emerge in a variety of different contexts. For example, the critical lure (e.g., *needle*) could be verbalized in association with several different list items (e.g., “*needle and thread*” and “*sharp point on the needle*”) or with novel semantic information (e.g., “stepped on a *needle* yesterday”).

Given our assumption that false recall is primarily a consequence of associations made between list items and critical lures generated during the presentation of list items, we developed the following hypotheses for the verbal reports. First, individuals who more frequently engage in semantic elaboration of list items (i.e., verbal reports reflect semantic elaboration) should be more likely to explicitly access and verbalize the critical lure during encoding than should individuals whose verbal reports primarily consist of verbatim rehearsal of list items. Second, verbalization of the critical lure during encoding would suggest the association of the lure to one or more of the list items and thus predict a higher probability of access to the lure (as reflected by its verbalization) during retrieval. Consequently, the likelihood of false recall of the lure should increase on the basis of the degree of semantic elaboration and number of different contexts created during encoding.

In our protocol experiment, there were 16 participants who thought aloud under the conditions of Experiment 1

and an additional 16 participants who thought aloud under the conditions of Experiment 2. The focus of our verbal-report experiment was to assess the relationship between encoding processes during study of list items and the likelihood of false recall. This analysis of the production of false recall stands in contrast to our manipulations in Experiments 1 and 2, which attempted to reduce the level of false recall across list structure conditions. Given that we were successful in replicating the results of our list structure manipulations across the first two experiments, our aim in the present experiment was only to provide data with which to examine variables that might predict generation of the critical lure during encoding and retrieval processes. Therefore, we were willing to accept that the estimated power to reproduce both of the effects of the experimental manipulation on false and correct recall was low (17% and 61%, respectively, for Experiment 1; 14% and 60%, respectively, for Experiment 2). Given that our predictions for the relation between verbalized encoding processes and false recall were equivalent for the methods used in Experiments 1 and 2, we combined the two data sets to increase power.

Method

Participants. A total of 32 introductory psychology students participated in the protocol experiment and received class credit for their participation. In contrast to the group administration used in Experiments 1 and 2, all participants were tested individually under the direct supervision of an experimenter.

Materials and Procedure. The materials and procedures were the same as in Experiments 1 and 2. Sixteen participants received the sequential presentation of list items as in Experiment 1, and 16 participants received the paired presentation of list items as in Experiment 2. The rate of presentation of words and other constraints on recall were identical to those of Experiments 1 and 2. All participants were instructed to “think aloud” or to verbalize their thoughts during the presentation of the word lists as well as during recall (Ericsson & Simon, 1993). More specifically, the participants were instructed:

As you are studying and recalling the list words, we would like for you to “think aloud” by vocalizing your thoughts. This process is quite similar to the experience of “talking to yourself” and saying everything that comes to mind as you work on the tasks. You should not attempt to explain your thinking, but rather to vocalize your thoughts as they occur to you. Try to do your best to simply verbalize your thoughts as they occur.

The participants were then given a series of warm-up tasks to acquaint them with thinking aloud. If the participants stopped talking during the warm-up tasks or during the experiment for more than 10 sec, they were instructed by the experimenter to “keep talking.” The participants’ verbal reports were tape-recorded and transcribed verbatim.

Two raters were presented with coding instructions and were asked to code each of the 320 protocols. The encoding and retrieval portions of each protocol were separated and coded blindly, such that the raters could not draw inferences when coding the retrieval section based on their knowledge of the encoding verbalizations (and vice versa). A third rater coded each of the protocols in the same manner, such that potential disagreements between the two initial raters could be resolved. All raters were blind to the nature and hypotheses of the experiment. Agreement rates between coders ranged from .89 to 1.0. Kappa coefficients, which assess proportionate agreement rates between coders by taking into account chance agreement, were within the *good* to *excellent* range of

between-coder agreement for all coding categories (Cooper & Hedges, 1994). Both agreement rates and Kappa coefficients for all coded variables are found in Appendix A.

Results

Because we successfully replicated our experimental findings across the first two experiments, recall performance data for sequential and paired presentations of list items were combined. Within-subjects ANOVAs, with list structure as a repeated measure, were used to assess both true and false recall performance (see Figure 1). Consistent with power calculations (36% for the combined sample of 32 participants), list structure failed to reliably influence false recall [$F(4,124) = 1.18, MS_e = 0.05, n.s.$]. However, the average frequency of false recall was 17%, which is not reliably different from that found in Robinson and Roediger’s (1997) 9-associates, 6-fillers condition.

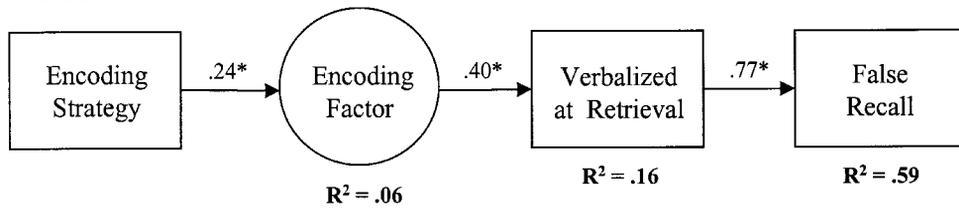
As expected from our power calculations for correct recall (99% for the combined sample), there was a significant main effect for list structure for correct recall [$F(4,124) = 8.70, MS_e = 0.007, p < .001$]. Consistent with Experiments 1 and 2, pairwise comparisons indicated that correct recall in the CU list structure ($M = .70, SD = .14$) was significantly higher than that in all other list structures ($ps < .01$). When the data from the CU list structure were excluded, there was no main effect for list structure [$F(3,93) = 1.43, n.s.$], and no further trend analyses were conducted.

Given that our analyses of the verbal reports focused on the production of false recall, the most important finding was that the proportion of false recall for the verbal report sample was roughly comparable to the average probability of false recall for Experiments 1 and 2 combined. Specifically, the average probabilities for false recall for the CU list structures were 7.8% for the verbal-report sample and 5.1% for Experiments 1 and 2 combined. For the other lists, the average probabilities of false recall were 14.5% for the verbal-report sample and 19.2% for Experiments 1 and 2 combined. Likewise, for correct recall, the corresponding probabilities of correct recall for the CU list structure and all other list structures were 60.1% and 69.8% for the verbal-report sample and 57.0% and 65.5% for Experiments 1 and 2 combined. Therefore, we further analyzed verbal reports collected for the relationships between encoding processes and the likelihood of false recall, given our experimental manipulations.

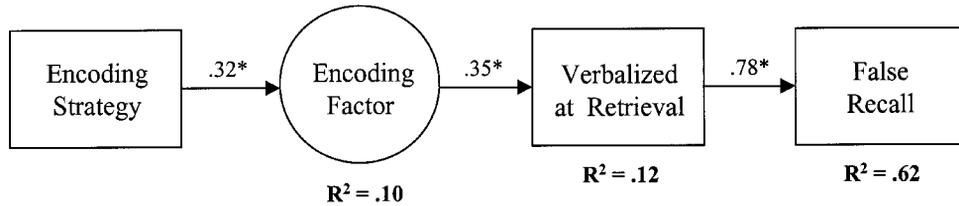
Path Models of Encoding Processes in False Recall

Sample verbal protocols for encoding and recall can be found in Appendix B. Two modeling approaches were taken to analyze the protocol data. First, we used a trial-based approach, in which we estimated relationships among coded variables across all list learning instances or across all list learning trials (320 list protocols; Model 1 in Figure 2). From our process model of false recall we hypothesized a sequential process in which one observ-

Model 1

* $p < .05$

Model 2

* $p < .05$

Model 2-R

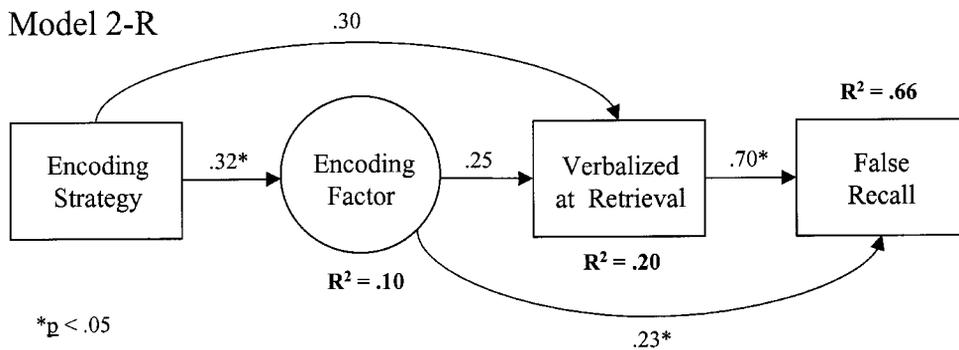
* $p < .05$

Figure 2. Path models of false recall based on verbal protocol data. Rectangular variables represent observed measures taken from the participants' verbal reports. Circular variables (encoding factor) represent the combined influence of the two encoding variables—namely, the number of unique contexts in which the lure was verbalized and whether or not the lure was verbalized during encoding.

able state, such as verbalization of the lure, would influence subsequent states and, eventually, the observable outcome of false recall. However, this trial-based approach did not control for the potential dependence between attempts by the same participants, as a path analysis treats each instance as an independent observation. Therefore, this analysis based on list-learning trials cannot be used to infer generalizability across a pattern of results to the population of participants. To adequately test for the generalizability of the relationship between the variables, we subsequently report the results of a second path model (Models 2 and 2-R in Figure 2), in which each participant contributes a single estimate for each variable—namely, the average across the 10 list-learning trials. In a final analysis, we assessed the merits of a process model by examining for a given participant whether the probability

of false recall on a given memory trial is higher when the same participant has verbalized the lure during encoding.

Our path models were developed using the following four coding variables: (1) type of encoding strategy, (2) the encoding factor, which combined the categories of whether or not the critical lure was verbalized during encoding and the number of unique contexts in which the critical lure was verbalized, (3) whether or not the critical lure was verbalized during recall, and (4) whether or not the critical lure was falsely recalled. Predictions were derived a priori on the basis of a simplistic process model of false recall. Specifically, use of a semantic elaboration strategy at encoding was predicted to increase the likelihood of the verbalization of the critical lure combined with the number of unique encoding contexts of the lure (encoding factor). Furthermore, the encoding factor was hy-

Table 2
Correlations (Trial-Based Approach) Among Verbal Protocol Coding Categories
Across Sequential and Paired Presentation of List Items ($n = 320$)

	Overall Encoding Strategy	Lure Verbalized at Encoding	Number of Contexts Lure Was Stated	Lure Verbalized at Retrieval	Lure Falsely Recalled
Overall					
Encoding Strategy		.262**	.239**	.172*	.077
Lure Verbalized at Encoding			.881***	.404***	.319**
Number of Contexts Lure was Stated				.430***	.381***
Lure Verbalized at Retrieval					.767***
Lure Falsely Recalled					
Encoding Factor	.236**	.960***	.956***	.402***	.349**

Note—"Encoding Factor" represents the combined influence of the "Lure Verbalized at Encoding" and the "Number of Contexts Lure Was Stated" coding categories. * $p < .05$. ** $p < .01$. *** $p < .001$.

pothesized to predict the formation of memory traces involving the critical lure and, consequently, verbalization of the lure at retrieval and subsequent false recall.²

Model 1: Trial-based approach. Pearson correlations for the trial-based approach ($n = 320$) are presented in Table 2. As displayed, correlations supported the previous hypotheses with regard to magnitude and appropriate direction. Model 1 was developed using the four coding variables to more directly assess the hypothesized relations between variables. The results of our instance-based approach to modeling the verbal protocol variables are displayed in Figure 2. The goodness-of-fit test, reported as a chi-square statistic,³ demonstrated good model fit to the data [$\chi^2(3) = 7.11, p = .07$]. The normed fit index (NFI) and the comparative fit index (CFI) also demonstrated good fit (.98 and .99, respectively). Furthermore, no standardized residuals exceeded the .10 criterion for concern. All model coefficients were statistically significant ($Zs > 1.96, p < .05$), and relationships between the variables were in the predicted direction. Specifically, the model suggests that when the participants used an encoding strategy that involved semantic elaboration of list items, they were more likely to verbalize it during encod-

ing and in a variety of different contexts. When the participants verbalized the lure at encoding, they were more likely to verbalize it at retrieval, which was associated with a higher level of false recall. In order to test the generalizability of this pattern of results, we analyzed the average probability of the outcomes for each participant.

Models 2/2-R: Participant-based approach. Pearson correlations for participant-based approaches are displayed in Table 3. Again, correlations were significant and in the appropriate direction, consistent with our original predictions. A second path model (Model 2) was used to more directly assess the hypothesized relations between variables. Figure 2 provides the results of the analysis of the independent data for the second path model. The goodness-of-fit test for this model demonstrated poor fit to the data [$\chi^2(3) = 8.56, p < .05$]. The NFI and CFI indices also demonstrated relatively poor fit (.81 and .86, respectively). Nevertheless, all of the model coefficients were statistically significant ($Zs > 1.96, ps < .05$), and relationships between the variables were in the predicted direction. Discrepancies between the observed and reproduced covariances resulted in two standardized residuals of .27 and .21, exceeding the .10 criterion for concern. A

Table 3
Correlations (Participant-Based Approach) Among Verbal Protocol Coding Categories
Across Sequential and Paired Presentation of List Items ($n = 32$)

	Overall Encoding Strategy	Lure Verbalized at Encoding	Number of Contexts Lure Was Stated	Lure Verbalized at Retrieval	Lure Falsely Recalled
Overall					
Encoding Strategy		.411*	.403*	.371*	.196
Lure Verbalized at Encoding			.954**	.458**	.472**
Number of Contexts Lure Was Stated				.473**	.528**
Lure Verbalized at Retrieval					.784**
Lure Falsely Recalled					
Encoding Factor	.360*	.930**	.941**	.348	.443*

Note—"Encoding Factor" represents the combined influence of the "Lure Verbalized at Encoding" and the "Number of Contexts Lure Was Stated" coding categories. * $p < .05$. ** $p < .01$.

revised model (Model 2-R in Figure 2) was constructed in which two paths were added, representing the large residuals in the initial model.

The goodness-of-fit test for Model 2-R resulted in a fail-to-reject decision, indicating good model fit [$\chi^2(1) = 1.44$, n.s.], and a significant improvement in fit when compared with the initial hypothesized model [$\chi^2(2) = 7.12$, $p < .05$]. The NFI and CFI indices also demonstrated good model fit (.97 and .99, respectively). Three of five of the model coefficients were statistically significant as noted in Figure 2 ($Z_s > 1.96$, $p < .05$), and relationships between the variables were in the predicted direction. The revised model suggests that the participants who frequently used an encoding strategy that involved semantic elaboration of list items were more likely to verbalize the critical lure during encoding. Most interestingly, the participants with a higher propensity to verbalize a critical lure at encoding were more likely to later exhibit false recall.

The success of the Model 2-R raises doubts about our proposed process model. If the participants were equally likely to falsely recall the lure on a memory trial whether or not they had verbalized the lure on that trial, then an account based on propensities for various activities would provide a more parsimonious account than our process model. To evaluate that possibility, we calculated phi coefficients between verbalization of lure on trial and false recall for each of the 19 participants who met the necessary conditions (i.e., having at least one instance of false recall and one instance of verbalization of lure). The average phi coefficient for these participants was .45, which was significantly greater than zero [$t(18) = 6.42$, $p < .001$]. The reliable relationship between verbalization of the lure and subsequent false recall within participants cannot be explained by differences between individuals and, thus, uniquely supports our process model.

Summary. From a detailed analysis of verbalized encoding processes, we hypothesized a process model (Model 1) that reliably predicts false recall in the DRM paradigm. A similar path model (Model 2-R) was found to predict the level of false recall for individuals on the basis of their frequencies of use of encoding strategies and verbalizations of the lure across trials. A final analysis found support for our process model by showing a reliable relation between verbalization of the lure and subsequent false recall on the same trials for each participant. The findings from our verbal reports are thus consistent with findings that participants' generation of specific semantic associates to specific list items will increase false memory (see Brainerd, Reyna, & Mojardin, 1999, for a review). In the following section, we will discuss the implications of our experimental manipulations and our path model for theories and approaches to the study of false recall.

DISCUSSION

In our study, two types of evidence demonstrate that encoding processes dramatically influence the likelihood

of false recall in the DRM paradigm. First, in an experimental manipulation of encoding context, false recall was virtually eliminated when list items were selected to induce encodings unrelated to the critical lure. Furthermore, this reduction in false recall was not an artifact of poor recall, because the recall of correct words from the CU list was reliably better than that of the standard DRM lists. Second, we collected concurrent verbal reports from a separate set of participants studying the same types of word lists to assess the cognitive processes during encoding that were associated with recall of the lure at test. We found that the participants verbally reported the critical lure during the encoding of the word lists, especially when the verbalized encodings of list items corresponded to phrases that contained semantic elaboration of list items. Overall, verbalization of the critical lure during encoding of a particular word list predicted an increased likelihood of false recall. We will first discuss our experimental findings in the context of current theories of false recall in the DRM paradigm, including Underwood's (1965) theory of implicit associative response (IAR) and fuzzy-trace theory (FTT; Reyna, 1998; Reyna & Brainerd, 1995; Reyna & Lloyd, 1998). Subsequently, we discuss the implications of our verbal report models in providing a more complete account of the processes underlying false recall.

Reduction of False Recall and Dissociation of True and False Recall

We found that false recall was dramatically reduced in lists with items that were organized to guide semantic encoding away from the critical lure. According to Underwood's (1965) IAR approach and other theories of spreading activation (e.g., Collins & Loftus, 1975), the same number of semantically related list items, regardless of organization within a list, should show the same level of activation of the critical lure and, thus, the same likelihood of false recall across list structures. In our experiments, our lists contained the same number of semantically related list items in all list structure conditions; however, we found an increase in false recall when semantic associates were more closely grouped together. Most interestingly, the CU list structure is structurally identical to the ones list structure, with the exception that the filler items are arranged randomly in the ones list structure, yet the likelihood of false is lower in the CU list structure. Mere association of semantically related list items, as proposed by spreading activation theories, does not fully account for our findings, and, instead, the contextual organization of list items influences the likelihood of false recall in the DRM paradigm.

One intriguing finding was that, within our CU list structure condition, we found the highest levels of correct recall and the lowest levels of false recall, suggesting a dissociation between true and false recall. When the encodings have been biased against such semantic relations, as in the CU list structure conditions, false recall is virtually eliminated (cf. Experiments 1 and 2). If this manipulation had reduced overall semantic processing of list

items, then fewer semantic connections to the critical lure and lower levels of false recall would occur, consistent with Togliani et al.'s (1999) notion of "more is less" in the DRM paradigm.⁴ However, our experiments show that this metaphor must be constrained to specific types of semantic processing. Specifically, when semantic processing is directed away from the lure or list theme, then false recall of the critical lure will be reduced, even though overall memory for list words is more accurate. The contextual organization of the CU list structure provides evidence for just this type of semantic processing; overall memory for list words is highly accurate, yet false recall of critical lures is diminished.

The dissociation between true and false recall observed in the CU list structure may be explained by FTT. FTT proposes that, when individuals encode information, they form two types of independent memory representations: verbatim representations (which include specific perceptual features) and gist representations (which include semantic information but lack perceptual details) (Reyna, 1998; Reyna & Brainerd, 1995; Reyna & Lloyd, 1998). In FTT, false recall of critical lures in the DRM paradigm is a consequence of dual-retrieval mechanisms. One retrieval mechanism involves direct access to verbatim traces of list items and is primarily responsible for correct recall. The other mechanism reconstructs items by processing gist representations and is primarily responsible for false recall.⁵ During the study of word lists from the DRM paradigm, semantic associations to the critical lure are generated and encoded into the gist representation. During subsequent recall of the list, the critical lure is retrieved by cues from the gist representation (Reyna, 1998; Reyna & Lloyd, 1998). FTT would predict that a gist representation of the DRM list theme is not formed in the CU list structure condition, and, therefore, verbatim traces of list items are accurately retrieved and the critical lure is not. However, this account is not particularly convincing, because FTT would also predict that activation of the gist representation would occur equally in all other list structures, leading to equally high levels of false recall. We found a decrease in false recall when semantic associates were spread throughout the list.

Clearly, the contextual organization of semantically associated list items dramatically influences the likelihood of both true and false recall. Our encoding manipulation demonstrates the importance of encoding processes in the DRM paradigm and has highlighted the need for current theories to account for the contextual influence of list items in false recall. To better understand the processes underlying false recall in the DRM paradigm, we also asked a set of participants to provide think-aloud protocols during the study and recall of word lists. We now turn to our verbal-report findings and their implications for theories of false recall in the DRM paradigm.

Verbal Reports and Theories of False Recall

Our introduction of verbal-report data on the encoding and retrieval of word lists has forced us to be more spe-

cific about the nature of different types of encoding processes and their relationship to false recall. According to our path models of false recall, during the presentation and study of list items, the generated encodings were more likely to have been semantically related to the critical lure when false recall was observed. Our encoding factor included whether or not the lure was stated during encoding and showed that the frequency of verbalization was directly related to the likelihood that the lure was falsely recalled. Below, we attempt to further examine our verbal protocols for the different types of semantic encodings generated (either automatically or deliberately) during encoding. In so doing, we believe that it may be possible to better characterize the gist representation proposed by FTT (Reyna, 1998; Reyna & Lloyd, 1998), as well as to assess the specific theoretical mechanisms that mediate access to the lure during encoding, such as with Underwood's (1965) IAR.

FTT describes encoding processes based on inferences from retrieval measures, whereas our verbal reports provide concurrent accounts of encoding and retrieval processes generated by participants in the DRM paradigm. With regard to FTT, our verbal reports provide an opportunity to explicate some of the particular processes and representations that reflect the formation and resultant cueing of the gist representation. For example, when the critical lure is verbalized during encoding, we can infer that the gist representation has been cued and memory traces involving associations to the lure are likely to have been formed. Furthermore, repeated verbalization of the lure during encoding may be considered repeated cueing or activation of the gist representation, which is hypothesized to increase false recall. This is consistent with our encoding factor and its ability to reliably predict verbalization of the lure at recall and false recall (Model 2-R).

Consistent with Collins and Loftus's (1975) theory of spreading activation, the critical lure is often related by mere association to an encoded word. For example, in the *chair* protocol described in Appendix B, the first verbalization of the lure *chair* may be retrieved due to word associates on the list (in particular, the list item *recliner*): "you can have a reclining CHAIR at your desk." Consistent with Underwood's (1965) IAR account, we can see the strategic and controlled use of the lure *chair* in the participant's rehearsal of list items and the lure: "sit, CHAIR, recliner." Likewise, a similar pattern may be observed in the *needle* protocol described in Appendix B: Automatic spreading activation of the presented list items *sewing*, *cook*, and *sharp* may have led to the activation of *needle* ("sharp is the NEEDLE"), followed by subsequent, deliberate use of the lure in combination with list items ("point on the NEEDLE" and "prick your finger with the NEEDLE that you are sewing").

Our verbal protocols contained instances in which the participants sometimes verbalized the critical lure in different semantic contexts during encoding. These instances are interesting theoretically for FTT, because they imply the existence of several gist representations. Dur-

ing retrieval, it may become difficult to distinguish which words and concepts had been generated during encoding and which had actually been presented as words on the list. We believe that this source confusion involving multiple, overlapping semantic encodings in memory is consistent with findings of increased levels of false recall with repeated memory testing (Roediger, 1996) and maintained levels of false recall after delayed testing (Payne & Elie, 1998). This finding might also account for the difficulties of attributing the source of a retrieved memory trace of the critical lure (Jacoby, Kelley, & Dywan, 1989; Johnson, Hashtroudi, & Lindsay, 1993).

Taken together, our verbal protocols appear to provide an insightful tracing of the processes underlying false recall. Although some support for the various theories may be observed across instances, the more formal path models have allowed us to explicate certain relationships as predictive of subsequent false recall. We believe that these relationships may help to clarify existing theories (i.e., FTT) by forcing theorists to provide a more detailed account of the proposed mechanisms underlying encoding and recall in the DRM paradigm. Our verbal reports also strongly suggest that the particular semantic encodings generated are quite variable and diverse across participants. We believe that future studies of the DRM paradigm will need to experimentally control encoding processes during the study of list items in order to further constrain subsequent retrieval processes and the likelihood of false recall. When experimental manipulations are coupled with process-tracing methods, such as verbal reports, our understanding of this complex phenomenon of false recall will be significantly advanced.

Future Directions

We believe that any theoretical framework that would be sufficient to explain the full range of findings relevant to the DRM paradigm will have to be quite complex and will have to accommodate many different types of encoding processes, as well as various kinds of retrieval and monitoring processes. Our manipulation of contextual encoding demonstrated unique effects both in reducing false recall and in illustrating a dissociation between true recall and false recall. By examining the verbally reported thoughts of the participants engaged in the DRM task, we have also attempted to identify more precisely several of the mechanisms involved in false recall. Despite our experimental manipulation that reduced the overall level of false recall and the limitation of the incompleteness of verbal reports (cf. Ericsson & Simon, 1993), it appears that various aspects of the aforementioned theories are at least partially consistent with verbal reports collected in our experiments. However, a complete model of the processes mediating false recall will require a more explicit account of the mechanisms involved, including an ability to generate testable predictions and to support recent findings in the DRM paradigm. Here, we have attempted to demonstrate the usefulness of coupling experimental and process-tracing

methods in critically examining current theories. Future research might benefit from furthering such an approach, leading to a more complete account of the encoding and retrieval processes underlying false recall.

REFERENCES

- BENTLER, P. M. (1995). *EQS: Structural equations program manual*. Los Angeles: BMPD Statistical Software.
- BRAINERD, C. J., REYNA, V. F., & MOJARDIN, A. H. (1999). Conjoint recognition. *Psychological Review*, **106**, 160-179.
- COLLINS, A. M., & LOFTUS, E. F. (1975). A spreading-activation theory of semantic memory. *Psychological Review*, **82**, 407-428.
- COOPER, H., & HEDGES, L. V. (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- DEESE, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, **83**, 17-22.
- ERICSSON, K. A., & SIMON, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.
- 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., HASHTRUUDI, S., & LINDSAY, S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- 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.
- NORMAN, K. A., & SCHACTER, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, **25**, 838-848.
- PAYNE, D. G., & ELIE, C. J. (1998, November). *Repeated list presentation reduces false memories for pictures and words*. Paper presented at the annual meeting of the Psychonomic Society, Dallas.
- 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.
- READ, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review*, **3**, 105-111.
- READER'S DIGEST EDITORIAL STAFF (Ed.) (1996). *Reader's Digest Oxford complete wordfinder*. Pleasantville: Reader's Digest Association, Inc.
- REYNA, V. F. (1998). Fuzzy-trace theory and false memory. In M. J. Intons-Peterson and D. L. Best (Eds.), *Challenges and controversies in applied cognition: Memory distortions and their prevention* (pp. 15-27). Mahwah, NJ: Erlbaum.
- REYNA, V. F., & BRAINERD, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning & Individual Differences*, **7**, 1-75.
- REYNA, V. F., & LLOYD, F. J. (1998, November). *Fuzzy-trace theory and source-monitoring explanations of false memories*. Paper presented at the annual meeting of the Psychonomic Society, Dallas.
- ROBINSON, K. J., & ROEDIGER, H. L., III (1997). Associative processes in false recall and false recognition. *Psychological Science*, **8**, 231-237.
- ROEDIGER, H. L., III (1996). Memory illusions. *Journal of Memory & Language*, **35**, 76-100.
- ROEDIGER, H. L., III, & MCDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.
- ROEDIGER, H. L., III, MCDERMOTT, K. B., & ROBINSON, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory: Volume II* (pp. 187-245). London: Psychology Press.
- 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.

- TOGLIA, M. P., NEUSCHATZ, J. S., & GOODWIN, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, *7*, 233-256.
- UNDERWOOD, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122-129.

NOTES

1. This procedure resulted in a 4 (random order of list structure) \times 4 (random order of list themes) \times 2 (within-subjects factor of two exposures of list structure) randomization schedule in which 2 participants were run in each cell of the design. Within each cell, the participants were randomly exposed to each list structure (eights, fours, twos, ones, and CU lists) on two occasions, resulting in 10 studied lists. Additionally, the schedule ensured that themed lists (e.g., *sleep*, *chair*, *mountain*, etc.) were randomly presented in each list structure across participants.

2. Due to the high degree of correlation between verbalization of the critical lure at encoding and the number of unique contexts generated ($r = .954, p < .01$), these variables were combined to yield a single factor score using principle components analysis. Although we believe these factors to be rather distinct constructs, their high degree of relatedness would have caused a degree of suppression or shifting in the model weights, thus masking their true relation with the outcome variables. The resultant factor score that combined the two variables

demonstrated a very similar pattern of correlation with the other variables in the model, including overall encoding strategy ($r = .360, p < .05$) and false recall ($r = .443, p < .05$).

3. In discussing the results of a path model, a goodness-of-fit test, reported as a chi-square statistic, is typically used to indicate how well the sample variance-covariance matrix is reproduced as a function of the parameter estimates of the proposed model. An obtained chi-square value with a probability level greater than a standard cut-off, such as .05, indicates adequate fit of the model to the sample. In contrast, a statistically significant chi-square value implies rejection of the model. Because the chi-square statistic is sensitive to sample size, we also report the results of two adjunct indicators of model fit: the normed fit index (NFI) and the comparative fit index (CFI). Both indicators can range from 0 to 1, with values of .90 or greater indicating good model fit (Bentler, 1995).

4. The "more is less" metaphor was used to illustrate the finding that conditions that increase the likelihood of correct recall (e.g., semantic processing and blocked presentation) simultaneously increase false recall. Thus, "more" studied and nonstudied items are remembered, such that overall remembering is "less" accurate.

5. It should be noted that because there are strong interitem semantic relations made between encoded list items in the DRM paradigm, later list items are likely to provoke retrieval of earlier list items, which may contribute greatly to the formation of the gist representation.

APPENDIX A

Protocol Coding Categories, Category Descriptions, Agreement Rates (AR), and Kappa Coefficients Observed for Sequential and Paired Presentations in the Verbal Protocol Experiment

	Description	AR		Kappa	
		Sequential	Paired	Sequential	Paired
Encoding					
Overall encoding strategy	A general assessment of the overall encoding strategy used by subjects, either <i>semantic elaboration</i> , in which subjects went beyond the literal presentation of list items, or <i>verbatim rehearsal</i> , which included only literal organization of list items and lures	.963	.937	.925	.872
Statement of lure	Yes or No (was the lure stated during encoding?)				
Overall encoding strategy of the lure	A general assessment of the encoding strategy of verbalizations of the lure, either <i>semantic elaboration</i> or <i>verbatim rehearsal</i>	.959	1.00	.958	1.00
Number of different contexts of lure (if mentioned more than once)	The number of different, unique encoding contexts in which the lure was stated	.889	1.00	.727	1.00
Retrieval					
Statement of lure	Yes or No (was lure stated during retrieval?)				
False recall of lure	Yes or No (was lure falsely recalled?)				

APPENDIX B

Verbal Protocol Samples

A typical protocol from the eights condition is provided below (italicized items during encoding are presented list words read by the participants from the computer, and italicized items during retrieval are recalled words as they were written by the participants).

Encoding

table I like to *sit* on top of tables I'm more comfortable tables have four *legs* you sit on them on table there are actual *seats* to sit on not at the table *couch* I like to sit on the couch it's more comfortable than the table you sit at the *desk* which is a type of table a *recliner* you can have a reclining CHAIR at your desk *sofa's* the same thing as a couch the *contents* of the desk which has a recliner my desk is very *tight* full of stuff I put my *hosiery* on top of the table hosiery *bel* these words don't relate I like to eat *potatoes* at my table sit CHAIR recliner on my desk there's a *computer* on my desk at the computer Jeff likes to look at the *football* scores on my *bed* my bed is right next to the desk

Here, we see that the participant verbalized the critical lure *chair* twice within the context of *recliner*. Interestingly, while this individual engages primarily in a semantic encoding strategy, she also engages in rehearsal of the lure.

Retrieval

table sit CHAIR *couch* *sofa* *bed* *desk* *recliner* *football* *computer* um on the table I like to sit on the table it was comfortable I like to sit on the CHAIR I prefer to sit on the couch than on the table another word for couch is a sofa next to my desk is the bed on the desk is the computer the desk has a recliner sometimes the desk is *tight* full of stuff on the computer my boyfriend likes to do his football scores um let's see what were the words that did not fit with this that were not part of the list that's about all I can remember sofa couch next to the desk is the bed on the desk is tight full of stuff

During recall, she falsely recalls *chair*, and does so with *sit*, one of the words that she rehearsed with *chair* during encoding.

The following protocol example is more representative of most encoding protocols in which semantic elaboration occurs and the critical lure is stated.

Encoding

thread *maneuver* *pin* and thread *head* a pin head maneuver and thread *eye* there is eyes on your head pin plus thread *ball* maneuver *sewing* pin, thread, ball oh gosh *cook* *sharp* sharp is the NEEDLE *shooter* I don't know *point* on the NEEDLE sewing with the *finger* *prick* your finger with the NEEDLE that you are sewing *thorn* prick your finger with a thorn use a *thimble* so you don't prick your finger with a thorn *collection* maneuver

In this encoding protocol, we again see the verbalization of the critical lure within a semantic elaboration encoding strategy. Again, *needle* is stated in a variety of different contexts (e.g., with *sharp*, *point*, and *prick*).

Retrieval

there's *maneuver* and there's a *thimble* so you don't have a *thorn* in you *finger* *sewing* with a NEEDLE and a *pin* maneuver and something else was sewing pin maneuver something else thimble for thorn prick I think sewing the NEEDLE with a pin and something more about sewing and sewing, NEEDLE, pin I don't know

In the corresponding retrieval protocol, *needle* is falsely recalled and is mentioned with related list items, but not necessarily with the same list items in which it was encoded.
