ACADEMIC PRESS

Journal of Memory and Language 47 (2002) 436–447

Journal of Memory and Language

www.academicpress.com

# The roles of associative responses at study and semantically guided recollection at test in false memory: the Kirkpatrick and Deese hypotheses

Steven M. Smith,* David R. Gerkens, Benton H. Pierce, and Hyun Choi

*Department of Psychology, Texas A&M University, College Station, TX 77843-4235, USA*

Received 11 July 2001; revision received 31 October 2001

**Abstract**

False recall is found for semantically related words that are not presented on both categorized and associatively structured study lists. Four experiments provide evidence that the associative list method produces false memories because of semantic processes involved in studying list words (the Kirkpatrick hypothesis), but that false memories produced by categorized lists occur because of the use of semantic knowledge at test (the Deese hypothesis). In a free association task, words from associative lists, but not categorized lists, tended to evoke critical words as responses, indicating that our categorized list words have low associative strength to critical nonpresented items. Studying those associative lists, but not the categorized ones, produced indirect priming effects in stem completion. Critical nonpresented words from categorized lists showed a priming effect only when participants were instructed at test to try to complete stems with studied list words (i.e., stem cued-recall). The results highlight important differences between categorized and associative list methods, and indicate that false memories can be caused by semantic processes that occur at the time of a memory test. 

*Keywords:* False memory; Indirect priming; Deese; Kirkpatrick

## False recall and stem completion: evidence of semantically guided recollection

People often misremember events, sometimes incorrectly claiming to recall or recognize non-presented words that are semantically related to studied items. For example, false recall and recognition of specifically targeted nonpresented words have been demonstrated frequently when participants study and recall lists constructed from words that are all associated with a single nonpresented critical word (e.g., Deese, 1959; Read, 1996; Roediger III & McDermott, 1995). Such lists, sometimes called DRM lists (named for popularly cited works by Deese, and by Roediger & McDermott), are often designed such that list words have high associative strength to critical nonpresented words, and will be referred to herein as *associatively structured lists*, or simply *associative lists*. A similar false recall effect occurs for categorized lists of words that omit highly typical or dominant category members (e.g., Smith, Gilliland, Gerkens, Pierce, & Tindell, 1998, 2000, 2001). That is, when highly dominant category

* Corresponding author.
*E-mail address:* sms@psyc.tamu.edu (S.M. Smith).

0749-596X/02/$ - see front matter 
PII: S0749-596X(02)00012-8

members (e.g., *chair* for the category *furniture*, *orange* for *fruit*) are left off of presented lists of common category members, the common nonpresented words are often falsely recalled. These false memories have been termed *semantic confusion errors* by Smith et al. (2001); such intrusions and false alarms are made because critical items are semantically related to the materials that were actually studied.

In the present study we focus on the question of when, in the course of learning and remembering, semantically influenced false memories occur. Are such semantic confusion errors caused by processes that occur as a function of studying word lists, processes that occur when materials are tested and remembered, or both?

In an early description of false memory effects, Kirkpatrick (1894) discussed the effects of semantically related material on false memories:

> There were some incidental illustrations of false recognition. About a week previously in experimenting upon mental imagery I had pronounced to the normal students ten common words. Many of these were recalled and placed with the memory list. Again, it appears that when '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 (Kirkpatrick, 1894, p. 608).

Kirkpatrick considered the effect of semantic confusion on false memory to be a study phenomenon, attributing the errors to the time when the list of words was first pronounced. Thus, the Kirkpatrick hypothesis states that false memories occur because of processes that take place when words are studied.

Some 65 years later, Deese (1959) found similar false memory effects using associatively structured lists, constructed such that each word on a list was closely associated with a single nonpresented linking word. For example, a list might contain the words "dream," "pillow," "nap," and "bed," in addition to other words that are strongly associated with the critical nonpresented linking word, "sleep." In contrast to Kirkpatrick's account, Deese described the effects of associated words as occurring at the time of the memory test, stating that, "in the process of recollection, words and concepts associated with remembered items will be added" (p. 21). The Deese hypothesis indicates that recall or recognition of some of the list words at test activates associations to the critical nonpresented word, thereby leading to false memory of the nonpresented words.[1] In the present study, we extend the Deese hypothesis to include effects at test of any type of semantic knowledge, including not only associations, but category knowledge and conceptual information, as well.

The preponderance of false memory research with associatively structured lists supports the Kirkpatrick hypothesis. Experiments that manipulate variables at study often find effects of those study factors on false recall and recognition. For example, some studies have reported effects of the level of processing of associative list words at study on subsequent false memory (e.g., Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toglia, Neuschatz, & Goodwin, 1999). Likewise, whether associative lists are presented in a blocked vs. random order at study affects false memory (e.g., Mather, Henkel, & Johnson, 1997; McDermott, 1996).

Indirect priming effects found by McDermott (1997) and McKone and Murphy (2000) are also more consistent with Kirkpatrick's (1894) hypothesis, supporting the notion that associative responses during study cause false memories. McDermott (1997) and McKone and Murphy (2000) found that nonpresented words from associatively structured lists were indirectly primed, as measured by a stem completion test. In those experiments few or no relevant associates to the critical nonpresented words were presented on the stem completion test, so associative responses leading to the critical word at test were unlikely to

[1] By contrasting these two hypotheses, we are not implying that they are mutually exclusive; we merely wish to give credit to those who first proposed the two hypotheses. As far as we know, neither Kirkpatrick nor Deese contradicted the alternative hypotheses, nor did they suggest that encoding and retrieval processes could not interact to produce the false memory effects they found. We have interpreted Deese's statement, "in the process of recollection, words and concepts associated with remembered items will be added," to mean that false memories can be due to processes that occur during testing, while participants are remembering word lists. Of course, Deese's main point was that the probability of producing an intrusion is "proportional to the average association strength of that item in the context of the material being recalled (p. 21)." Deese's ideas about association strength could be used to explain false memory effects that are due to processes that occur during initial study as well as during the test.

occur. Thus, the stem completion findings in those experiments can be attributable to processes that occur during study. Further evidence of such associative responses at study causing false memories has been reviewed and provided by McEvoy, Nelson, and Komatsu (1999). Whether such associative responses are cases of unconscious spreading activation, implicit associative responses (IARs, e.g., Bousfield, Whitmarsh, & Danick, 1958; Kimble, 1968; Underwood, 1965) that are covertly but consciously experienced, or strengthening of pre-existing associations (Zeelenberg, Shiffrin, & Raaijmakers, 1999) is not directly addressed in the present study. What is addressed is whether semantic knowledge causes false memories solely because of study processes, or whether effects can also occur because of processes that take place at the time when a test occurs.

Little evidence from studies with associative list materials supports the Deese hypothesis, which attributes false memories to processes initiated when memory testing occurs, and some results directly contradict the Deese hypothesis. For example, Roediger III, McDermott, and Marsh (2000), using associative list materials, tested critical nonpresented words that were preceded by varying numbers of associated list words. The Deese hypothesis predicts that false recognition and recall should be increasingly likely as more list words are encountered at test. Roediger et al. found, however, that the test position of critical nonpresented words did not affect false memories on a word stem cued-recall test, or on a yes/no recognition test. These results directly contradict the Deese hypothesis that false memories are test-induced.

The likelihood that people would have associative responses to critical nonpresented words seems quite high for associatively structured lists because those list words were selected on the basis of high backward association strength. Backward association strength refers to the probability that a presented list item will evoke the critical nonpresented word (e.g., *dream* evokes *sleep*; Robinson & Roediger III, 1997). Associative responses that activate the critical nonpresented item, therefore, are more likely for lists with high backward association strengths. The categorized lists used by Smith et al. (1998), on the other hand, were not chosen to selectively evoke associative responses of critical nonpresented category members. That is, although critical nonpresented words are conceptually related to presented words in the categorized list method, the categorized lists may have low backward association strengths in relation to critical nonpresented items.

Despite the low backwards associative strengths of some categorized lists, such lists might evoke false memories for reasons other than associative responses that occur at study. The category structure of categorized lists may guide memory at the time of testing, more in line with the Deese hypothesis. Such a hypothesis is consistent with the findings of Smith et al. (2000), who found that the category structure of categorized lists strongly influenced false recall of critical nonpresented category members. In that study, nonpresented category members that were highest in output dominance and typicality were more likely than nondominant atypical nonpresented category members to be falsely recalled. Therefore, although studying such list words might not activate critical nonpresented words (the Kirkpatrick hypothesis) via backwards associations, recalling such lists should nonetheless elicit critical items (the Deese hypothesis) via the category structure.

In the present study we first selected a set of categorized and associative lists, and compared the associative strengths of the two types of lists. In Experiment 1 participants gave free associations to words from a set of associative and categorized lists. It was predicted that free association to words from associatively structured lists would be far more likely than members of categorized lists to elicit critical nonpresented words as responses. In Experiments 2 and 3, it was predicted that indirect priming of nonpresented categorized words would not occur, because such effects appear to be due to associative processes at study. It was also hypothesized that category knowledge used to guide recollection would cause false recall of critical nonpresented items from categorized lists.

Experiment 4 tested a different prediction of the Deese hypothesis with categorized list materials. In Experiment 4, instructions to complete stems were either incidental, as in Experiments 2 and 3 (i.e., participants were told, "fill in each stem with any word that comes to mind"), or intentional (i.e., "use studied words to fill in stems whenever possible"). Intentional instructions on the stem completion test of Experiment 4 should highlight the category structures of the categorized lists; therefore such instructions were predicted to bring about an indirect priming effect in stem completion.[2] That is, intentional instructions change the task into a stem cued-recall test that should show the same retrieval-guided false recall effect seen in the other recall tests. Because the instructions were manipulated at the time of the test it was predicted that an effect would occur with categorized list materials.

## Experiment 1

Two sets of lists were selected: associative lists drawn from McDermott's (1997) materials, and categorized lists taken from Smith et al.'s (2001) materials. In Experiment 1, participants were asked to give free associations to both types of list words, with the 10 words in each list presented in a single block. Critical items from the lists, which were not presented on the lists, were expected to occur as associative responses quite often for associatively structured lists, but far less often for categorized lists.

### *Method*

#### *Participants*

The participants in all of the reported experiments were undergraduate volunteers who completed part of a course requirement by their participation. Participants were recruited for group sessions using posted signup sheets. Volunteers could enroll for any of many experiments, including the present ones. There were unequal numbers in the treatment groups because unequal numbers of participants enrolled for different experimental sessions. Each session was held in a group of approximately 5–15 participants at a time. Forty-two undergraduate students participated in Experiment 1.

#### *Materials*

Four 10-item categorized lists and four 10-item associative lists were presented to participants. The categorized lists were modified (shortened to match the length of the associative lists) from those used by Smith et al. (2001). The associative lists were drawn from the ones used by McDermott (1997). The critical items used in the present experiments are listed in Appendix A. Copies of the four associative and the four categorized lists, with blank spaces next to each item, were provided for the participants. Type of list alternated on the page. All 10 words in each list were given as a single block of words.

#### *Procedure*

Participants were instructed to read each word and then write the first word that came to mind in the blank next to the word. If they thought of a word they had previously used, participants were instructed to use the word again. Participants worked at their own pace until they had provided one associate for every list item. Approximately half the subjects started with a categorized list and approximately half started with an associative list. There was no effect of counterbalancing order, so the reported results are collapsed across this variable.

### *Results*

A significance level of $p < .05$ was used on all statistical tests for all experiments reported, unless otherwise specified. The number of critical free associates was compared for associatively structured lists vs. categorically structured lists, using the number of times per list a critical word was given as a free associate to list members as the dependent variable. There was a significant effect of list type [$t(41) = 8.01$, $SE = .18$]; associative lists (1.57) evoked critical items as free associates more than 10 times as often as did categorized lists (.14).

### *Discussion*

As expected, it was found in Experiment 1 that critical words from associatively structured lists were given as responses in the free association task at a very high rate, about one-and-a-half times per 10-word list. Those associatively structured word lists were designed for the purpose of evoking the targeted linking words as associates, and the results of Experiment 1 confirm that the words had

[2] In the intentional instruction condition, the stem completion task becomes a stem cued-recall test. The finding that studying a categorized word list leads to more stem completions that use critical nonpresented words in the intentional condition is clearly due, in large part, to intentional, rather than unintentional memory processes, and technically should not be termed an "indirect priming" effect. We retained the use of the term "indirect priming" in the intentional condition, however, to make such effects more clearly comparable with effects of the same manipulation (i.e., prior study of the related categorized word list) in the incidental condition, not only in Experiment 4, but in previous experiments.

exactly the desired effect. Critical words from categorically structured lists, however, were not frequently given as responses on the free association test. Words from the categorized lists evoked critical items at a rate of less than one-tenth that found for the associatively structured list words, approximately once for every eight categorized lists. These results indicate that when participants study categorized lists of words, they are very unlikely to experience critical nonpresented category members as associative responses.

## Experiment 2

The intention of Experiment 2 was to retest McDermott's (1997) finding of indirect priming in stem completion with both associative and categorized lists, using several procedures designed to emphasize the perceptual nature of the test, minimizing recollection and conceptually driven processing at retrieval. In Experiment 2, the test stimuli on the word stem completion test were presented at a fast rate, only 1 s per word stem, a rate fast enough to minimize deliberate conceptually driven retrieval (Weldon, 1993). Furthermore, no critical words were ever presented on the study lists. This was done to minimize participants' attempts to recollect list words to complete word stems. In fact, in the present experiments no word stems could be completed by any studied list words. It was expected that these precautions would prevent, or at least minimize the occurrence of deliberate conceptually driven retrieval processes on the word stem completion test.

In Experiment 2 of the present study many participants were recruited for both associative and categorized list conditions to ensure enough statistical power to detect even a small effect.

### *Method*

#### *Participants*

There were 122 volunteers who participated in Experiment 2. None had been participants in Experiment 1.

#### *Materials*

The same eight modified lists (four 10-word associative lists and four 10-word categorized lists) used in Experiment 1 were again used in Experiment 2. Participants studied four of the eight lists, including two categorized and two associative lists. The words on each list were presented at a rate of 1.5 s per word. All 10 words for each list were presented together in a single block. Each list was preceded by a screen denoting which list was to follow (e.g., "List 1" was displayed for 2 s prior to the first list). Participants were provided with blank forms on which to respond to the various tasks during the experiment. In addition to the list presentation, a letter counting task, a number counting task, a stem completion task, and a free recall task were all presented on a large television screen. All tasks were created using Microsoft PowerPoint.

#### *Design*

List type (associative vs. categorized) was manipulated within-subjects, as was priming (primed, unprimed). Priming was counterbalanced between-subjects, such that each list was primed (studied) in half of the treatment conditions and unprimed (not studied) in the other conditions.

#### *Procedure*

Participants were instructed to watch the word lists that were presented on a monitor, and to try to remember the words in each list. Participants were instructed not to write anything down during the list presentation. After list presentation, participants were told that prior to the memory test, they were to complete a series of other unrelated tasks. The first was a number counting task. In this task, a series of seven-character strings (letters and numbers) were presented for 1 s each with a 3-s interval between string presentations. Participants counted the number of digits within each string during presentation and recorded the number during the 3-s interval. Next was the stem completion task. Each of 28 word stems was presented for 1 s, and participants had 5 s to write down a word to complete the stem. Eight of the 28 stems corresponded to the critical items. Because of the counterbalancing conditions of list presentation, each critical stem was indirectly primed in two conditions and unprimed in two conditions. The stem completion score from the unprimed condition was used as the baseline completion rate for that stem. Participants were instructed to complete stems with the first word that came to mind. No mention of the previous lists was made, and none of the stems could be completed by list items. The third task was the letter counting task. In this task participants saw a series of letter arrays for 2 s each with a 10-s interval. Each array was composed of six rows of seven characters each. Participants counted how many times the letter 'N' appeared in the array.

Finally, participants had a free recall test. Participants were instructed to recall as many of the original list items as they could. They were also instructed to try to keep items from the same list together by making a separate column for each list. Participants were instructed not to guess, but to include only items that they remembered from the lists.

### Results

Indirect priming was tested using a 2 (primed, unprimed) $\times$ 2 (associative, category list type) repeated measures design with critical item stem completion as the dependent variable. Scores were collapsed across counterbalancings for the present analyses. There was an indirect priming main effect [$F(1, 121) = 10.52$, $MSE = .11$] and a list type main effect [$F(1, 121) = 193.17$, $MSE = .06$]. However, these main effects were qualified by a significant interaction [$F(1, 121) = 8.68$, $MSE = .10$]. Table 1 shows how critical stems for the associative lists were more likely to be completed by the critical word than were stems matching the critical nonpresented category members. Furthermore, follow-up simple main effect analyses revealed that there was an indirect priming effect only for the associative list critical items [$F(1, 121) = 12.37$, $MSE = .17$] and [$F < 1$] for the associative and category items, respectively.

The indirect priming effect size in the associative list condition was $f = .32$, a medium effect size (see Cohen, 1988). The power for detecting an indirect priming effect of the same magnitude as that found in the categorized list condition was .99 in Experiment 2. The nonsignificant stem completion indirect priming effect size was $f = .04$ in the categorized list condition.

Table 1
Proportion of stems completed by critical items for associative and categorized lists as a function of priming in Experiment 2

| Priming | List type | |
|---|---|---|
| | Categorical | Associative |
| Primed | .10 (.02) | .50 (.03) |
| Not primed | .09 (.02) | .31 (.03) |
| Indirect priming effect | .01 | .19 |

*Note*. Standard errors appear in parentheses.

Table 2
Mean proportions of critical intrusions and correct recall for Experiment 2

| Measure | List type | |
|---|---|---|
| | Categorical | Associative |
| Critical intrusions | .36 (.03) | .72 (.03) |
| Correct recall | .40 (.02) | .39 (.02) |

*Note*. Standard errors appear in parentheses.

Concerning the free recall data, a paired $t$ test revealed that there was also a significant difference in proportion of false recall of critical items across list type [$t(122) = 8.15$, $SE = .04$], with greater false recall of nonpresented words from associative lists. This difference was only for critical words; the number of items correctly recalled and the number of noncritical intrusions did not differ across list type (Table 2).

### Discussion

A statistically reliable indirect priming effect in word stem completion was found using associatively structured lists, replicating results by McDermott (1997) and McKone and Murphy (2000). In spite of precautions to eliminate deliberate recollection on the stem completion test, the presentation of associatively structured lists caused a medium-sized indirect priming effect on the primarily perceptually driven test. This result indicates that the influence of associative processes on these semantic confusion errors occurs at study, consistent with the Kirkpatrick hypothesis. In contrast, studying categorized word lists did not affect stem completion of critical nonpresented words, even though critical words from categorized lists were falsely recalled quite often. Our failure to find a stem completion indirect priming effect with categorized list materials does not appear to be due to a lack of statistical power; the type 2 error rate for that effect was calculated as less than .01. Thus, consistent with the Deese hypothesis, the results show that the categorized lists caused false memories on a test that emphasized recollection (recall), but not on a test in which recollection was minimized.

The results of Experiment 1 indicate that while studying categorized lists of words, participants are not likely to experience critical nonpresented category members as associative responses. In sharp contrast, studying lists that have high backwards associative strength frequently should evoke associative responses corresponding to critical nonpresented words. This difference may explain why, in Experiment 2, indirect priming occurred in stem completion for associative lists, but not for categorized lists.

## Experiment 3

In Experiments 1 and 2 it was shown that associative and categorized lists both evoke false memories, but that associative lists have high levels of associative responses and significant indirect priming effects, whereas categorized lists rarely evoke critical items as associative responses, and do not cause indirect priming effects. These differences in indirect priming effects might be attributed to qualitative differences in the list materials. Unfortunately, in Experiment 2 materials were used that produced different levels of false memories for associative vs. categorized lists. In Experiment 2 associative list items were falsely recalled twice as often as were critical nonpresented items from categorized lists. This difference might have occurred for a number of reasons, including the fact that critical associative list items often had been used to complete word stems prior to the recall test. In the primed condition critical words from associative lists were used in critical stem completions 50% of the time, and even in the unprimed condition, 31% of the critical stems were completed with critical nonpresented words. Thus, recall of critical words from associative lists was contaminated and possibly inflated by an extra exposure to critical words on the stem completion test. The baseline and primed levels of stem completion for words from categorized lists were only 9% and 10% for the unprimed and primed conditions, respectively. Overall, the levels of false recall were very high for both categorized (.36) and associative (.72) lists. Are indirect priming effects in stem completion affected by the type of list (associative vs. categorized), or are they simply caused by lists that evoke strong false memory effects?

In Experiment 3 measures were taken to use associative and categorized list materials that would evoke equivalent levels of false memories. This equivalence was accomplished in part by selecting a subset of the categorized and associative lists from Experiment 2 that had similar levels of false memories. In addition, the targeted selection of items resulted in baseline (unprimed) stem completion rates that were more equivalent for the associative and categorized list conditions.

Half of the critical word stems in Experiment 3 corresponded to associative lists and half corresponded to categorized lists. The critical items were either unprimed, indirectly primed (i.e., the corresponding list was studied), or directly primed (i.e., critical words were presented in an incidental word rating task). It was predicted that direct priming, relative to the unprimed condition (used for the base rate), would produce more stem completions of critical items from both associative and categorized lists. An indirect priming effect, however, was predicted to occur only for critical items from associative lists. A final recall test over all of the studied lists was given as a manipulation check to assure that critical items corresponding to both associative and categorized lists had equivalent levels of false recall. Such a check is necessary to show that differences between the associative and categorical list conditions in indirect priming is not simply due to differences in false memory.

### *Method*

#### *Participants*

There were 118 volunteers who participated in Experiment 3.

#### *Materials*

A set of 24 unrelated words was used for a pleasantness rating task. In the direct priming condition, four of the words from this task were the critical items. The same four critical word stems (corresponding to associative lists for critical nonpresented items *anger* and *music*, and corresponding to categorized lists for critical items *pants* and *orange*) were used in each of the three priming conditions.

Two 10-word categorized lists and two 10-word associative lists were shown to participants, as described in Experiment 2. The four critical lists, corresponding to the four critical word stems, were seen only in the indirect priming condition. In the unprimed and direct priming conditions, four noncritical lists (two associative and two categorized lists) were the study lists. Each participant studied only two lists of each type, with list type alternated.

#### *Procedure*

In each condition participants had five tasks: (1) pleasantness rating, (2) list learning, (3) number counting, (4) stem completion, and (5) free recall of the studied lists of words. In the pleasantness rating task participants rated each of 24 unrelated words according to how the word made them feel, ranging from −3 (very bad) to +3 (very good). Three seconds were given for each word rating. The procedures for tasks (2)–(5) were the same as described for Experiment 2.

*Design*

Type of list (associative vs. categorized) was manipulated within-subjects, as in Experiment 2. Type of priming (unprimed vs. indirect priming vs. direct priming) was a between-subjects factor. In the unprimed condition the four critical word stems did not correspond to any words on the incidental pleasantness rating task, nor to any of the four studied lists of words. In the indirect priming condition the critical word stems did not correspond to words on the incidental task, but they were critical (nonpresented) items from the four studied lists. In the direct priming condition the four critical stems corresponded to four critical words that were included on the incidental pleasantness rating task, but they did not correspond to the studied lists of words.

*Results*

A paired $t$ test was calculated comparing the proportion of falsely recalled critical items for the associative vs. categorized lists. The proportions of false recall did not significantly differ across list type [$t(31) = 1.44$, $SE = .065$, $p = .16$]. The mean proportion false recall rate for critical associative list items was .56 ($SE = .037$), as compared with .66 ($SE = .065$) for critical categorized list items.

A $2 \times 3$ (priming × list type) ANOVA was conducted using proportion of stem completions as the dependent measure, with list type (associative vs. categorized) as a within-subjects variable and priming (unprimed, indirectly primed, or directly primed) as a between-subjects variable. There was a significant main effect of priming [$F(2, 115) = 3.29$, $MSE = .072$] and a list type main effect [$F(1, 115) = 14.00$, $MSE = .097$]. However, these effects were qualified by a marginally significant list type × priming interaction [$F(2, 115) = 2.91$, $MSE = .097$, $p = .058$] . There were different patterns of priming for the two list types. Tukey tests showed that stem completion for the indirect priming condition was significantly greater than the unprimed condition for associative lists [$t(73) = 1.99$, $p = .05$], but not for categorized lists [$t(73) = -1.58$, $p = .12$]. In fact, for categorized list items, the mean stem completion rate was slightly, but not significantly less in the indirectly primed condition than the unprimed condition. The mean stem completion rates for Experiment 3 are shown in Table 3. A Tukey test revealed that stem completion was significantly greater for directly primed items than for unprimed items for both list types combined [$t(84) = 2.73$, $SE = .038$]. The base rate of stem completion in the unprimed conditions was equivalent for associative (.17) and categorized list items (.12).

Table 3
Mean stem completion rates on critical stems for categorized and associative lists as a function of priming in Experiment 3

| List type | Priming condition | | |
|---|---|---|---|
| | Direct | Indirect | Unprimed |
| Categorized | .20 (.041) | .05 (.026) | .12 (.033) |
| Associative | .30 (.047) | .34 (.079) | .17 (.044) |
| *n* | 43 | 32 | 43 |

*Note*. Standard errors appear in parentheses.

*Discussion*

The free recall test in Experiment 3 indicated that false recall levels for associative and categorized lists were equivalent, alleviating the interpretive confusion from Experiment 2, in which false recall levels differed for the two types of lists. The baseline (unprimed) stem completion rates for associative and categorized list items were also equivalent. Replicating the results of Experiment 2 were the indirect priming effects, which were found to be significant for critical items from the associative lists in Experiment 3, but not for critical items from the categorized lists. These results show that the difference in indirect priming effects for the two types of list materials cannot be attributed to differences in false memory strength.

Extending beyond the effects of Experiment 2 are the direct priming effects found in Experiment 3, which were found for both types of materials. These direct priming effects cannot be labeled indirect priming effects because the critical words were actually seen in the experiment (albeit, not on the study lists). The direct priming level for associative list items was equivalent in magnitude to the indirect priming effect for those items. This finding is consistent with the notion that studying associative lists may produce perceptual-like experiences of critical nonpresented words during study, whereas studying categorized lists of words does not.

## Experiment 4

The results of Experiments 2 and 3 (and other findings with the same materials by Smith et al., 1998) show that false memories in the categorized list method occur in direct memory tests, such as cued-recall, but not on indirect memory tests that are primarily perceptually driven, such as a speeded stem completion test. With the categorized lists used in the present study, false memories appear to be a byproduct of processes involved in deliberate recollection, consistent with the Deese hypothesis. In Experiment 4 we further tested the Deese hypothesis by manipulating the instructions given at test. Participants studied three categorized word lists in anticipation of a recall test. Following presentation of the lists, but before the final category cued-recall test (used in lieu of the free-recall tests in Experiments 2 and 3), participants were given a word stem completion test. The test included stems of the critical nonpresented category members from the three studied lists (primed condition), plus stems of critical items from three nonstudied categorized lists (unprimed condition). Two instruction conditions were used, manipulated between-subjects. In the incidental condition participants were told to complete each stem with the first word that came to mind. In the intentional instruction condition, participants were told to complete stems with words from the studied lists whenever it was possible. Thus, the intentional condition was essentially a test of stem cued-recall. The six critical word stems could not be completed by any of the presented words.

It was predicted that critical nonpresented category members would be falsely recalled on the final cued-recall tests in all conditions. No indirect priming effect in stem completion was predicted to occur in the incidental condition, consistent with the results of Experiments 2 and 3. In the incidental condition of Experiment 4, as in Experiments 2 and 3, the stem completion test was designed to minimize recollection. Therefore, no indirect priming effect was expected for the incidental condition. In the intentional condition, however, participants were encouraged to try to recollect list words as they completed the word stems. Thus, the Deese hypothesis predicted that an indirect priming[2] effect would be found in the intentional condition of the stem completion test, because deliberate recollection would be affected at test by the same semantic processes that can lead to false memories.

As in Experiments 2 and 3, many participants were recruited for Experiment 4 to mitigate questions about statistical power of the experiment for detecting an effect.

### *Method*

#### *Participants*

There were 261 participants in Experiment 4.

#### *Design and materials*

Six 15-item categorized lists from Smith et al. (2001) were used in Experiment 4 (no associative lists were used). Participants studied three of the six lists. In each of two counterbalancing conditions, nonpresented items from three lists were indirectly primed and three were unprimed. The same number counting and letter counting tasks from Experiment 2 were used. A similar stem completion test was also used. A final cued-recall test was given, with category names serving as list cues.

#### *Procedure*

The procedure was the same as that used in Experiment 2, with the exception that participants only viewed three lists (instead of four). Cued-recall tests were given sequentially over each of three lists after the stem completion test was done. On the stem completion test, half of the participants received the incidental instructions (the same as described for the stem completion tasks in Experiments 2 and 3, i.e., "complete the stem with the first word that comes to mind"). The other half were instructed to complete stems with studied list words whenever possible (intentional instruction), but to use the first word that came to mind if they could not remember a list word that completed a stem.

### *Results*

A 2 (primed vs. unprimed) $\times$ 2 (intentional vs. incidental instruction) mixed ANOVA using critical item stem completion as the dependent variable was calculated to examine the influence of intentional recollection in this task. Instruction was a between-subjects factor and priming was a within-subjects factor. The results are collapsed across counterbalancing conditions. There was a significant indirect priming main effect

Table 4
Proportion of critical stems completed with critical nonpresented items as a function of priming and instruction in Experiment 4

| Priming | Instruction | |
|---|---|---|
| | Intentional | Incidental |
| Primed | .75 (.07) | .28 (.05) |
| Unprimed | .31 (.04) | .25 (.04) |
| Indirect priming effect | .44 | .03 |
| *n* | 143 | 118 |

*Note.* Standard errors appear in parentheses.

[$F(1,259) = 20.18$, $MSE = .35$] and a significant instruction main effect [$F(1,259) = 25.51$, $MSE = .35$]. These main effects were qualified by a significant priming × instruction interaction [$F(1,259) = 16.02$, $MSE = .35$]. Simple main effects analyses for each instruction reveal that an indirect priming effect occurred in the intentional instruction condition [$F(1,143) = 31.42$, $MSE = .44$], but not in the incidental instruction condition [$F < 1$]. Table 4 shows the proportions of critical stem completions for each condition.

The effect size of the indirect priming effect was $f = .47$ in the intentional condition, but only $f = .04$ in the incidental condition. The power for detecting a medium sized stem completion indirect priming effect in Experiment 4 was .99.

Independent *t* tests failed to reveal differences across instruction condition in proportion of critical intrusions [$t(259) = 1.57$], correct recall [$t(259) = .55$] or number of noncritical intrusions [$t(259) = .28$].

### Discussion

Once again, there was no indirect priming effect in stem completion for nonstudied category members when participants were instructed simply to complete each word stem with the first word that came to mind. Although the same words were falsely recalled nearly half of the time, the indirect priming effect was negligible in the incidental condition. Once again, the lack of an effect was not due to a lack of statistical power, which was calculated as .99 in Experiment 4.

There was a powerful indirect priming effect when participants were told to deliberately complete word stems with words that had been on the studied lists. More than twice as many stems were completed with critical nonpresented items when corresponding category members had been studied. Ironically, the critical words participants used to complete stems had not been on the studied lists; thus, the instruction to use *studied* words increased the likelihood of using *nonpresented* words on the stem completion test. The effect of instruction in Experiment 4 was quite strong (Cohen's $f = .47$) and stands in contrast to the small effects of instructions on indirect priming in McKone and Murphy's (2000) experiments. This difference between associative and categorized list materials again highlights an important difference between the two methods, consistent with the idea that false memories with associative list materials are more due to processes that occur at study, whereas categorized lists produce false memories because of processes that come into play during retrieval.

## General discussion

A set of categorized lists of words was used in three experiments to demonstrate false recall of nonpresented common category members, consistent with findings by Smith et al. (2000) and Smith et al. (2001). Are these false memories caused by semantic processes that occur at study (the Kirkpatrick hypothesis), or at test (the Deese hypothesis)? Our consistent failure to find indirect priming of common nonpresented category members in Experiments 2–4 indicates that studying those categorized lists did not activate the critical nonpresented items. However, we did find clear test effects in the categorized list method. In Experiment 4 it was shown that, although indirect priming did not occur for categorized lists when incidental instructions were given on the stem completion test (i.e., "complete word stems with the first word that comes to mind"), there was a large effect in the intentional instruction condition, when deliberate recollection was encouraged on the stem completion test (i.e., "complete word stems with studied list words whenever possible"). Thus, the manipulation of instructions at test had a potent effect on priming of nonpresented category members. Collectively, the results of the present experiments with categorized list recall provide clear evidence in support of the Deese hypothesis that false memories can result from semantic confusion that occurs when memory is tested.

Results from the categorized lists used in Experiments 2–4 stand in contrast with findings from associative list materials. Although both types of materials produced false recall, only study of associative list materials caused indirect priming of nonpresented words (Experiments 2 and 3). Failures to find indirect priming of nonpresented common category members were not due to a lack of statistical power, which was very high in Experiments 2–4. Nor were those failures due to the use of a stem completion test that was insensitive to priming; incidental presentations of critical category members (i.e., direct priming of critical words) in Experiment 3 led to a significant priming effect on the stem completion test. Furthermore, the possibility that such indirect priming effects did not occur because such effects depend upon greater "false memory strength" was rejected by the results of Experiment 3. In that experiment, critical items from categorized lists were falsely recalled at least as often as critical items from associative lists, yet indirect priming occurred only for associative list materials.

What accounts for the observed differences between associative and categorized list methods in terms of indirect priming effects? Although the present experiments can offer no definitive answer to this question, the evidence is consistent with the idea that high backwards associative strength between list words and critical nonpresented items is the key. The likelihood that list words will evoke associations to critical words during study is far greater in the associative lists than in the categorized lists used in the present study, as shown by Experiment 1. It should be pointed out that, in theory, categorized lists can be high in backwards associative strength; the lists used in the present study, however, were not.

The indirect priming effects with associative lists (Experiments 2 and 3) replicate results of McDermott (1997) and McKone and Murphy (2000). Indirect priming effects in implicit stem completion are not due to associative responses at test, because few (or no) associates were ever presented on stem completion tests in these studies. Furthermore, studies with associative lists show effects of study manipulations on false memory (e.g., Mather et al., 1997; McDermott, 1996; Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toglia et al., 1999), but no effects of test manipulations (e.g., Roediger III et al., 2000). As a whole, false memory research with associative lists supports the Kirkpatrick hypothesis, that false memories result from semantic processes that occur at study.

The pattern of results with associative lists is consistent with the theory that associated responses activated when list words are first presented cause subsequent false memories, as originally suggested by Kirkpatrick (1894). Critical nonpresented words are often given as free associates to words on associative lists, but not categorized lists, a finding confirmed by Experiment 1. Underwood (1983) stated, "... it is assumed that subjects may produce IARs to the perceived words, and that these IARs are those that would occur if the subjects were given word-association tests (p. 133)." Following from this point of view, that the patterns of free associates to list words indicate like patterns of IARs, it seems reasonable that indirect priming in stem completion may depend on IARs generated at study. For example, reality monitoring errors are cases in which memories of events that were imagined are confused with memories of events one has perceived (Johnson & Raye, 1981). IARs may be conceived as covert, but conscious events. An unconscious activation interpretation provides a different explanation. This theory states that the presentation of list members unconsciously activates a single critical item, which is then falsely remembered because of its heightened state of activation (e.g., Ayers & Reder, 1998; McEvoy et al., 1999). Although the present experiments do not critically test these two explanations, it should be noted that both are consistent with the Kirkpatrick hypothesis.

The support for the Deese hypothesis provided by the present experiments has important implications for false memory research that seeks to separate study from test effects. Categorized lists, such as those used in the present study, appear to elicit false memories at the time of testing without also involving the study effects found in the associative list method. The categorized list evidence in support of the Deese hypothesis is also relevant to applied situations involving suggestion or implication at test. In applied settings, recollection might be influenced by semantic knowledge, such as category knowledge or conceptual information; therefore, one might conceivably elicit false memory because of test factors.

## Acknowledgments

We are very thankful to Tommy Roshek, Robin Maltsberger, Jason Martini, Michele Monies, Tracy Lockless, Jessica Williamson, Shannon

Creekmur, and Dave Morton for their assistance in running the reported experiments.

## Appendix A. Critical items from associative and categorized lists

| Categorized list items | Associative list items |
|---|---|
| Chair | Anger[a] |
| Football | Music[a] |
| Knife | Needle |
| Pants[a] | River |
| Orange[a] | Smoke |
| Robin | Spider |

[a] Indicates items used in Experiment 3 to equate false levels.

## References


Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: predictions from an activation-based memory model. *Psychonomic Bulletin & Review, 5*, 1–21.

Bousfield, W. A., Whitmarsh, G. A., & Danick, J. J. (1958). Partial response entities in verbal generalization. *Psychological Reports, 4*, 703–713.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Deese, J. (1959). On the predictions of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67–85.

Kimble, G. A. (1968). Mediating associations. *Journal of Experimental Psychology, 76*, 263–266.

Kirkpatrick, E. A. (1894). An experimental study of memory. *Psychological Review, 1*, 602–609.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: remember/know judgments and memory characteristics questionnaire compared. *Memory Cognition, 25*, 826–837.

McDermott, K. B. (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.

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, and Cognition, 25*, 1177–1194.

McKone, E., & Murphy, B. (2000). Implicit false memory: effects of modality and multiple study presentations on long-lived semantic priming. *Journal of Memory and Language, 43*, 89–109.

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.

Rhodes, M. G., & Anastasi, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin & Review, 7*, 158–162.

Robinson, K. J., & Roediger III, H. L. (1997). Associative processes in false recall and false recognition. *Psychological Science, 8*, 231–237.

Roediger III, H. L., & McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 803–814.

Roediger III, H. L., McDermott, K. B., & Marsh, E. J. (2000). The role of retrieval factors in development of false memories. In *Presented at the 41st annual meeting of the psychonomic society, New Orleans, November*.

Smith, S. M., Gilliland, T. R., Gerkens, D. P., Pierce, B. H., & Tindell, D. R. (1998). Dissociations of false memory measures: cued-recall vs. stem completion. To be presented at the convention of the Psychonomic Society, Dallas, TX.

Smith, S. M., Tindell, D. R., Pierce, B. H., Gilliland, T. R., & Gerkens, D. R. (2001). Source memory failure in episodic confusion errors. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 362–374.

Smith, S. M., Ward, T. B., Tindell, D. R., Sifonis, C. M., & Wilkenfeld, M. J. (2000). Category structure and created memories. *Memory & Cognition, 28*, 386–395.

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., 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.

Underwood, B. J. (1983). *Attributes of memory*. Glenview, IL: Scott, Foresman and Company.

Weldon, M. S. (1993). The time course of perceptual and conceptual contributions to word fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1010–1023.

Zeelenberg, R., Shiffrin, R. M., & Raaijmakers, J. G. W. (1999). Priming in a free association task as a function of association directionality. *Memory and Cognition, 27*, 956–961.