False memories are hard to inhibit: Differential effects of directed forgetting on accurate and false recall in the DRM procedure

John G. Seamon; Chun R. Luo; Elizabeth P. Shulman; Sarah K. Toner; Selin Caglar

Wesleyan University, Connecticut, USA.

To cite this Article: Seamon, John G., Luo, Chun R., Shulman, Elizabeth P., Toner, Sarah K. and Caglar, Selin (2002) 'False memories are hard to inhibit: Differential effects of directed forgetting on accurate and false recall in the DRM procedure', Memory, 10:4, 225 — 237

To link to this article: DOI: 10.1080/09658210143000344
URL: http://dx.doi.org/10.1080/09658210143000344
False memories are hard to inhibit: Differential effects of directed forgetting on accurate and false recall in the DRM procedure

John G. Seamon, Chun, R. Luo, Elizabeth P. Shulman, Sarah K. Toner, and Selin Caglar

Wesleyan University, Connecticut, USA

Directed forgetting research shows that people can inhibit the retrieval of words that they were previously instructed to forget. The present research applied the directed forgetting procedure to the Deese/Roediger and McDermott (DRM) recall task to determine if directed forgetting instructions have similar or different effects on accurate and false memory. After studying lists of semantically related words, some participants were told to forget those lists, whereas other participants were not. All participants were then shown additional lists to remember. Following study, all participants were asked to free recall as many of the studied words as possible, including those they were previously instructed to forget. Directed forgetting instructions inhibited the accurate recall of studied words, but not the false recall of nonstudied critical words, whether measured by a within-participant or between-participants design. Contrary to an implicit activation hypothesis, false memories survived instructions to forget. These findings were reviewed in terms of fuzzy trace theory and the activation/monitoring approach to false memory.

In an immediate free recall task, Deese (1959) presented participants with lists of semantic associates that converged on nonstudied words. For example, for the list consisting of thread, pin, eye, sewing, sharp, point, pricked, thimble, haystack, pain, hurt, and injection, the converging associate was needle. Deese found that the participants often erroneously recalled the nonstudied converging associates (hereafter referred to as critical words). Nearly four decades later, Roediger and McDermott (1995) revived interest in this memory illusion. In their first experiment, these researchers replicated Deese’s finding that participants often erroneously recall nonstudied converging associates (hereafter referred to as critical words). In their second experiment, Roediger and McDermott tested memory for the same word lists by recognition, and they added a meta-memory judgement. For any recognised word, the participants had to make a remember/know judgement by stating whether they specifically remembered the word’s occurrence at study or they merely knew the word had been presented. Roediger and McDermott found that critical words were falsely recognised as frequently as studied words, and these false recognitions were often characterised as remembered in the remember–know meta-memory task.

Thus, following exposure to semantically related word lists, Roediger and McDermott found that participants not only falsely recalled or recognised critical words that were never presented, but participants also indicated that they remembered some aspect of the prior occurrence of those critical words at study. These remember judgements for nonstudied critical words suggested that participants were capable of “mentally re-experiencing” events that had never actually occurred (Roediger & McDermott, 1995).
basis of Deese’s and Roediger and McDermott’s research, this approach to the study of false memory has come to be called the Deese/Roediger and McDermott (DRM) procedure.

Following the publication of Roediger and McDermott’s (1995) results, numerous researchers have manipulated a wide variety of subject and methodological variables to determine their effect on false recollection in the DRM procedure (e.g., McDermott, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996; Read, 1966; Schacter, Verfaellie, & Pradere, 1996; Seamon, Luo, & Gallo, 1998). However, despite intensive research on variables that can attenuate or enhance false memory (for reviews, see Kopelman, 1999; Roediger, McDermott, & Robinson, 1998; and Schacter, Norman, & Koutstaal, 1998), relatively few studies have examined the ability of participants to inhibit their false memories in this procedure.

THE INHIBITION OF FALSE MEMORY

Studies of the ability of participants to edit memory in general and inhibit false memory in particular are relatively rare. Yet the ability to edit memory is critically important in conditions where memory accuracy is essential. For example, when courtroom witnesses are asked to “tell the whole truth and nothing but the truth”, they are explicitly asked to edit their recall (Koriat, 2000). Within the DRM procedure, a number of studies have sought to determine if participants could edit their memory performance so as to reduce or eliminate their susceptibility to the false memory effect. Two different approaches to memory editing have been employed. The first approach involves attempts to attenuate the effect by forewarning instructions. For example, Gallo, Roberts, and Seamon (1997) gave one group of participants instructions that included forewarning of the false memory effect by a demonstration of the effect, a demonstration of a typical DRM word list, and a description of the specific false memory effect, prior to study and test. Compared to other groups of participants who were either not forewarned or merely urged to minimise false alarms, the forewarned group made fewer false recognitions of critical words than the other groups, yet all participants, including the forewarned, still made numerous false recognitions. Subsequent research by McDermott and Roediger (1998) and Gallo, Roediger, and McDermott (2001) has replicated and extended this basic finding: When participants are given multiple DRM word lists to study, they have a difficult time editing their false memory at test, even when they understand this memory illusion.

The second way to study editing processes in the DRM procedure is less direct than forewarning. It involves presenting participants with the same word lists for multiple study and test trials. Presumably, when participants are repeated shown the same lists for study and test, they have an ever-greater opportunity to learn the specific contents of each list. Thus, accurate memory for studied words should increase over trials, whereas false memory for critical words should decrease. However, in studies involving five study–test trials for the same word lists, false memory was not eliminated. McDermott (1996) found that correct recall of studied words increased substantially over trials, whereas false recall of critical words was diminished, but not eliminated. Similar results were obtained by Schacter, Verfaellie, Anes, and Racine (1998) when memory was tested by recognition, and by Seamon et al. (2001b), who combined multiple study–test trials with forewarning instructions and still observed a false recognition effect.

Together, these studies involving forewarning and multiple study–test trials are consistent in demonstrating that participants are able to edit their memory performance in the DRM procedure. However, these studies also show that this editing ability is limited. Whether tested by recall or recognition, all of the studies reported only limited success in suppressing false memories. False memories in this procedure are difficult to inhibit. One novel approach that has heretofore not been attempted is to try to inhibit false memories by the directed forgetting procedure.

MEMORY INHIBITION BY DIRECTED FORGETTING

In the directed forgetting procedure, participants are presented with a set of words and instructions to remember or forget each word. In the item method, after each word is presented for study, an explicit cue to remember or forget that word is provided. In the list method, only one cue to remember or forget is provided, normally after half of the words in a list are presented (Basden, Basden, & Gargano, 1993; MacLeod, 1999). Results from three decades of research with the directed forgetting procedure indicate that when
participants are asked to recall all of the studied words, they recall more words that they were trying to remember than those they were instructed to forget (for recent reviews, see Bjork, Bjork, & Anderson, 1998; Johnson, 1994; MacLeod, 1998).

To explain the directed forgetting effect, researchers have suggested differential rehearsal or retrieval inhibition interpretations, depending on whether the item or list method is employed. Because a directed forgetting effect is observed for both recall and recognition with the item method, retrieval mechanisms are thought to play little part in this process. Directed forgetting may occur because the to-be-remembered items are rehearsed more than the to-be-forgotten items at study. However, with the list method, a directed forgetting effect is typically found with recall, but not recognition, implicating retrieval, rather than rehearsal differences (see Basden et al., 1993, for a review). To explain the list method results, researchers have suggested a retrieval inhibition interpretation: The to-be-forgotten information is inhibited and does not interfere with information to-be-remembered. Yet the strength of the to-be-forgotten information in memory is unaffected because when retention is tested by recognition instead of recall, memory for the to-be-forgotten information is unimpaired (e.g., Elmes, Adams, & Roediger, 1970; Geiselman, Bjork, & Fishman, 1983).

In their discussion of retrieval inhibition, Bjork and Bjork (1996) have argued that memory updating is an adaptive cognitive process, as cues to forget render prior learning out-of-date, and retrieval inhibition makes this material less retrievable. If so, we wondered what would be the fate of false memories for critical words from lists that were subjected to directed forgetting instructions in the DRM procedure. Would these false memories be inhibited by instructions to forget in the same manner as accurate memories, or might these false memories once again demonstrate their robustness and survive, even when their semantically associated studied words were inhibited?

THEORETICAL IMPLICATIONS OF DIRECTED FORGETTING

To determine the effect of directed forgetting instructions on accurate and false recall, we used the list method of directed forgetting because the DRM words are organised by lists. Moreover, because previous research has shown that a reliable directed forgetting effect is obtained for recognition memory only when the item method is used (e.g., Basden et al., 1993), we tested memory by free recall instead of recognition. Under intentional learning conditions, participants were presented with either 8 or 12 DRM lists for study. In effect, we conducted two experiments with the directed forgetting procedure, one with 8 lists, and one with 12 lists. We varied the numbers of lists for study to determine if this variable influenced the results and, more important, to provide an internal reliability test of our findings. After half of the lists were presented, some participants were instructed to forget the lists they had just ‘‘erroneously’’ studied and remember the remaining lists (forget/remember instructions). Other participants merely paused after the first half of the lists and received no forgetting instructions (remember/remember instructions). After all of the lists were presented, all participants were asked to free recall as many of the studied words as possible, including, for the directed forgetting participants, those words that they were previously instructed to forget.

Based on previous research, the results from the directed forgetting procedure can be analysed in different ways, depending on the type of baseline conditions employed (see Bjork et al., 1998; MacLeod, 1998). First, a within-participant comparison can be made for those participants instructed to forget the lists after half of them were presented. Accurate and false recall can be compared for the first half and the second half of the lists for these forget/remember participants. According to MacLeod (1998), this within-participant method has come to be the preferred measure of directed forgetting. Second, a between-participants measure can also be employed when a remember/remember baseline condition is added. In this case, accurate and false recall can be compared for the first half of the lists by comparing participants who were instructed to forget these lists with participants who were trying to remember them. The present experiment permitted both measures of directed forgetting and, thus, provides another reliability check for our findings. Regardless of measure, accurate recall of studied words and false recall of critical words for both the to-be-forgotten and to-be-remembered lists were compared to determine if directed forgetting instructions had similar or different effects on accurate and false memory.
Because previous research has shown that directed forgetting instructions inhibit the recall of studied words (see Bjork et al., 1998; Johnson, 1994; MacLeod, 1998), the main focus of the present research was on the effect of these instructions on false memory and the theoretical implications of the results. Prior findings with the DRM procedure have been interpreted in terms of different theoretical positions, and new findings based on the effect of directed forgetting instructions may help differentiate these positions. Two fundamentally different approaches are found in Underwood’s (1965) implicit activation response hypothesis and Brainerd and Reyna’s (1996, 1998) fuzzy trace theory.

The implicit activation response hypothesis

Several researchers have previously used Underwood’s (1965) implicit activation response hypothesis to interpret their findings in the DRM procedure (e.g., Robinson & Roediger, 1997; Roediger & McDermott, 1995; Seamon et al., 1998). According to Underwood, when participants encode words during study in a memory task, those words can activate their semantic associates. This activation may occur automatically, with the participants covertly verbalising the semantic associates as the list items are presented. This hypothesis suggests that when participants study DRM word lists (e.g., thread, pin, eye, . . .), the participants can activate representations for nonstudied words (e.g., needle) because those words are the highest semantic associates of the list items. Participants may subsequently falsely remember those words on the basis of that prior activation.

Roediger et al. (1998) suggested that a source monitoring approach (e.g., Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981) can provide a similar account of false memory. Participants may activate representations for critical words during study and later confuse the source of these internally activated representations with those that were activated by the externally presented word lists, resulting in a misattribution error for the critical words. For the present research, the important point about both the implicit activation and source monitoring approaches is that participants are hypothesised to treat critical words as studied words and remember both in a comparable manner. Underwood (1965, p. 122) was quite specific on this point by stating that an implicit activation response is not a hypothetical construct. Rather, he said, from the participant’s perspective, it represents a stimulus that actually occurred. Accordingly, without additional assumptions, the implicit activation and source monitoring approaches would suggest that directed forgetting instructions should operate in a similar manner for both accurate and false memory in the DRM procedure. Because previous research has shown that directed forgetting instructions inhibit accurate memory for studied words, these instructions should also inhibit false memory for critical words, if critical words are encoded in the same manner as studied words.¹

Fuzzy trace theory

According to fuzzy trace theory, memory judgements are based on verbatim or gist traces that are stored at the time of study. Verbatim traces represent the surface details of physical stimuli, whereas gist traces represent the meaning or theme of the stimuli (see Brainerd & Reyna, 1996, 1998; Brainerd, Reyna, & Mojardin, 1999, 2001). In the DRM procedure, researchers have suggested that accurate recall of studied words is driven largely by verbatim traces, whereas false recall of critical words is based predominantly on gist traces that correspond to list themes (e.g., Payne et al., 1996; Seamon et al., 2001a; Toglia, Neuschatz, & Goodwin, 1999). Whereas the implicit activation and source monitoring approaches treat studied and critical words in the

¹ An alternative prediction, derived from Underwood’s hypothesis and suggested by Charles Brainerd (personal communication), is that directed forgetting instructions could have a greater inhibitory effect on critical words than studied words. If participants covertly verbalise words as the lists are presented, the participants may activate traces for both semantically associated studied words, as well as semantically associated nonstudied words (critical words). This could produce greater activation for explicitly presented studied words than only implicitly activated critical words, resulting in greater resistance to retrieval inhibition (i.e., less directed forgetting) for list members than critical words. For present purposes, however, we will follow the lead of previous researchers (e.g., Roediger & McDermott, 1995; Seamon et al., 1998) and adopt a more restricted interpretation of Underwood’s hypothesis by assuming that implicit activation responses apply only to the most common semantic associate of the list members (the critical word). This interpretation yields the more cautious prediction that both list words and critical words can be activated and subject to retrieval inhibition from directed forgetting instructions.
same fashion, fuzzy trace theory holds that the representations for studied words and critical words are different. Moreover, this difference has been used by fuzzy trace theorists to account for dissociations of experimental variables on accurate and false memory in the DRM procedure (see Roediger et al., 1998, for a review of these variables).

Fuzzy trace theory is actually an opponent processes theory. Memory judgements in the DRM task are based on a generation process that produces gist traces at study that lead to false memory, as well as a memory editing process that monitors retrieval and can reduce false memory. For example, when participants are given multiple study–test trials in the DRM procedure, list repetition at study increases the availability of verbatim traces for those words, and thereby increases the discriminability between studied and critical words for participants editing their retrieval during test. At some point, participants realise that a critical word is highly similar to previously studied words, but different and not a list member. Thus, accurate memory increases over trials, whereas false memory decreases (McDermott, 1996; Schacter et al., 1998; Seamon et al., 2001b). Brainerd et al. (2001) refer to this process of discriminating accurate from false memories as recollection rejection. Yet the research on multiple study–test trials also shows that false memories are robust and resist elimination. In fact, recent research by Seamon et al. (in press) demonstrated that false memory, although diminished, remained substantial in the DRM procedure even after the study lists were repeated 25 times. Together, these studies indicate that participants have only a limited ability to edit their false memory in this task.

Because false memories can be experimentally dissociated from accurate memories (see Roediger et al., 1998), the directed forgetting procedure provides a novel dissociation test. Unlike the implicit activation or source monitoring approaches, fuzzy trace theory suggests that directed forgetting instructions will effect accurate recall more than false recall. Specifically, if the reduction in false memory following forewarning or multiple study–test trials can be attributed to a limited ability of participants to suppress false memory by recollection rejection, this same editing mechanism may be used to inhibit the false recall of critical words following instructions to forget. Directed forgetting instructions should inhibit the recall of studied words, as previous research has already demonstrated (e.g., Bjork et al., 1998; Johnson, 1994; MacLeod, 1998). But, because participants have only a limited ability to edit their false memory, these instructions may have only a modest effect, if any, on inhibiting the recall of critical words. False memories for critical words may even survive instructions to forget.

In summary, without additional assumptions, the implicit activation and source monitoring approaches predict that directed forgetting instructions will inhibit accurate and false recall in a similar manner in the DRM procedure. Fuzzy trace theory, on the other hand, predicts that directed forgetting instructions will inhibit accurate recall of list words more than false recall of critical words. This experiment was conducted to test these alternatives.

**METHOD**

**Participants**

The participants were 140 Wesleyan University students, approximately 17 to 22 years old, who served as paid volunteers or received introductory psychology credit. None had taken part in any related memory research. Of these participants, 80 saw 8 word lists at study, and 60 saw 12 word lists. In the 8-list condition, 32 participants received remember/remember instructions, whereas 48 participants received forget/remember instructions. In the 12-list condition, 20 participants received remember/remember instructions, whereas 40 participants received forget/remember instructions. Given the importance of a potential null finding on the effect of forget instructions on false recall, more participants were tested with forget/remember than remember/remember instructions in each list condition in order to enhance the power of this within-participant (forget/remember) measure of directed forgetting.

**Materials**

We used two sets of 16 or 24 DRM word lists obtained from Roediger and McDermott (1995). Each list was composed of 15 words that were all converging associates of a nonstudied critical word. Within each list, the order of the words was constant, and the strongest associates to the critical word normally occurred first. Each stimulus set of 16 or 24 lists was counterbalanced in two ways. First, we divided each set of 16 or 24 lists
into two subsets of 8 or 12 lists. Thus, there were two subsets of 8 lists (labelled A and B), and two subsets of 12 lists (also labelled A and B). Regardless of the number of lists shown at study, half of the participants in each condition studied the appropriate Set A and half studied the corresponding Set B. Second, we further divided each of these list subsets into two list order subsets, such that half of the participants saw one subset of the lists first (Lists 1–4 or Lists 1–6), whereas the remaining participants saw the second subset of lists first (Lists 5–8 or Lists 7–12). This second counterbalancing manipulation distributed any potential list effect across the within-participant forget/remember conditions.

### Procedure

At the start of the experiment, we told the participants that they would see lists of words on a computer screen, followed by a free recall test of those words. Using standard intentional learning instructions, we asked the participants to remember as many words as possible for immediate recall. Our instructions included a practice set of four unrelated words and a written recall test. Following the instructions, the participants were presented with 8 or 12 lists of 15 words in blocked fashion on an Apple computer. The words were typed in 24 point, upper case, black, Helvetica font, and they were presented sequentially in the centre of the screen at a rate of 2 seconds per word with no delay between words. Each list was separated by a red, 1-second “Next List” prompt. All of the participants were tested individually, and they were not informed about the semantic relatedness of the words in the study lists.

After half of the study lists were presented (the first four lists in the 8-list condition or the first six lists in the 12-list condition), hereafter referred to as Block 1 of the study lists, the word “PAUSE” appeared in red on the screen. For participants in the remember/remember condition, the experimenter entered the test room and told the participant that the computer stopped to provide a brief rest period during study and that more word lists would follow shortly. Approximately 1 minute later, the remaining lists (the second four lists in the 8-list condition or the second six lists in the 12-list condition), hereafter referred to as Block 2, were presented.

For participants in the forget/remember condition, the experimenter also appeared after the first half of the lists had been presented and the computer had stopped. But, now, instead of describing the pause as a brief rest period, the experimenter expressed surprise that the computer had paused. She said that she had made a procedural error and presented the wrong set of lists for study. The experimenter apologised and asked if the participant would stay a bit longer to study the correct set of lists. All participants agreed to complete the experiment, and the “correct” set of lists was then presented in the same manner as before. During this pause of approximately 1 minute, all participants were specifically asked to forget the words they had erroneously just studied (i.e., the first four or six lists of Block 1) to avoid confusing those “incorrect” word lists with the “correct” word lists (i.e., the second four or six lists of Block 2) they were about to study. For obvious reasons, this forget/remember procedure is sometimes referred to as the “whoops” procedure (Bjork et al., 1998).

After all of the study lists were presented, all participants were given a free recall test in which they were asked to write down, in any order, as many of the previously shown words that they remembered, regardless of prior recall instructions. Participants in the forget/remember condition were informed that they were not really given any incorrect lists, and the experimenter apologised for deceiving them. These participants were specifically instructed to recall all of the words that they could remember, including those words they were previously asked to forget. No time limit was placed on the free recall. Following recall, all participants were fully debriefed so that they understood the necessity for the use of deception.

### RESULTS

The results were analysed for accurate recall of studied words, false recall of nonstudied critical words, and false recall of unrelated words (intrusions). Two different types of analyses were conducted because the effects of directed forgetting instructions could be appropriately analysed by both within-participant and between-participant designs. To measure directed forgetting in a within-participant design, only the results from participants in the forget/remember condition were analysed. For this design, number of lists at study (8 or 12), list set (A or B), and list order within set (1 or 2) were between-participants
variables, whereas list block (1 or 2) was a within-participant variable. To measure directed forgetting in a between-participants design, only the results from Block 1 for all participants were analysed. For this design, study instructions (remember/remember or forget/remember), number of lists at study (8 or 12), list set (A or B), and list order within set (1 or 2) were all between-participants variables. Because stimulus list set and list order within list set did not systematically affect performance for any dependent measure in either design, these variables will not be considered further in the results.

The goal of this research was to determine the effect of directed forgetting instructions on accurate and false recall in the DRM procedure. Table 1 shows the results in terms of mean recall proportions for studied words and critical words for 8 lists, 12 lists, and both list conditions pooled. As previous research has repeatedly shown, the DRM procedure produces levels of false recall that are comparable to accurate recall (e.g., Read, 1996; Roediger & McDermott, 1995). More important for the present study, regardless of whether directed forgetting instructions were examined as a within-participant variable (Block 1 vs 2 of the forget/remember condition) or as a between-participants variable (remember/remember vs forget/remember conditions for Block 1), the same results were obtained. Table 1 readily shows that directed forgetting instructions inhibited accurate, not false, recall. This finding was supported by the results of different analyses of variance done separately on studied and critical words.

Accurate recall of studied words

The top two rows of Table 1 show the mean proportion of studied words that participants recalled correctly as a function of their instructions, the number of study lists, and list block. When directed forgetting instructions were examined as a within-participant variable, an analysis of variance for the forget/remember condition showed that correct recall was greater in Block 2 (remember) than Block 1 (forget), \( F(1, 84) = 35.24, MSe = .27, p < .0001 \), recall was greater following 8 lists than 12 lists, \( F(1, 84) = 8.59, MSe = .11, p < .005 \), and there was an interaction of list block and number of lists, \( F(1, 84) = 9.36, MSe = .07, p < .005 \). An analysis of the interaction showed that there was no difference in the recall of words from Block 1 for 8 or 12 lists, \( F < 1 \), but there was more recall from Block 2 for 8 lists than 12 lists, \( F(1, 84) = 15.07, MSe = .12, p < .001 \). More important, more words were recalled from Block 2 than Block 1 for both the 8-list, \( F(1, 84) = 44.50, MSe = .01, p < .001 \), and 12-list \( F(1, 84) = 3.80, MSe = .01, p < .06 \), conditions.

Because directed forgetting instructions are confounded with list blocks in a within-participant design, it is important to demonstrate an effect of directed forgetting when the data are examined in a between-participants design. When directed forgetting instructions were examined as a between-participants variable, using the results from Block 1, the results showed only a main effect of instructions. Participants instructed to remember the words from Block 1 recalled more studied words than participants who were initially

### TABLE 1

<table>
<thead>
<tr>
<th>Number of study lists</th>
<th>8-lists block</th>
<th>12-lists block</th>
<th>Pooled lists block</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word type and instructions</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Studied words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remember/Remember</td>
<td>.26</td>
<td>.29</td>
<td>.21</td>
</tr>
<tr>
<td>Forget/Remember</td>
<td>.19</td>
<td>.31</td>
<td>.18</td>
</tr>
<tr>
<td>Critical words</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remember/Remember</td>
<td>.27</td>
<td>.27</td>
<td>.25</td>
</tr>
<tr>
<td>Forget/Remember</td>
<td>.28</td>
<td>.30</td>
<td>.24</td>
</tr>
</tbody>
</table>

Block 1 refers to the first half of the study lists (the first 4 or 6 lists presented); Block 2 refers to the second half of the lists (the remaining 4 or 6 lists). Remember/Remember participants were initially instructed to remember words from Blocks 1 and 2; Forget/Remember participants were initially instructed to forget words from Block 1, but remember those from Block 2. At test, all participants were instructed to free recall words from both blocks, regardless of prior instructions.
instructed to forget them, $F(1, 132) = 6.96$, $MSe = .08$, $p < .01$. There was no main effect of number of lists and no interaction of instructions and number of lists, both $ps > .12$.

In summary, as shown in Table 1, initial instructions to forget lists of words, whether analysed as a within-participant or between-participants variable, significantly inhibited the subsequent free recall of those words. Directed forgetting instructions reduce accurate recall by approximately 20% by either analysis, and the magnitude of this reduction is comparable to other studies of directed forgetting using the list method with recall (e.g., Basden et al., 1993).

**False recall of critical words**

The bottom two rows of Table 1 show the mean proportion of critical words that participants falsely recalled as a function of their instructions, the number of study lists, and list block. The same analyses that were applied to the accurate recall results were repeated on the false recall data. When directed forgetting instructions were examined as a within-participant variable, an analysis of variance showed no effect of blocks, number of lists, or their interaction, all $ps > .22$. Similarly, when directed forgetting instructions were examined as a between-participants variable on the results Block 1 alone, the same findings were obtained. There was no effect of instructions, number of lists, or their interaction, all $ps > .44$. Unlike the recall of studied words, these analyses, as indicated by the results in Table 1, demonstrated that directed forgetting instructions, whether analysed as a within-participant or between-participants variable, had no measurable effect the false recall of critical words.

Our finding that directed forgetting instructions dissociated accurate and false recall is new and theoretically important, but it is based on the failure of the false recall data to reject the null hypothesis. To ensure that our experiment had enough statistical power to detect a possible directed forgetting effect on false recall, we performed separate power analyses for the within-participant and between-participants designs. Because the mean recall proportions for studied words and critical words were highly similar in both designs, our measures of effect size for false recall were based on the corresponding results for accurate recall of studied words for each design. The effect size was large (.54) for the within-participant design and small to medium (.36) for the between-participants design. The results, based on the GPOWER general power analysis program (Erdfelder, Faul, & Buchner, 1996), indicated the power to be .99 for the within-participant design and .65 for the between-participants design. Because the within-participant design is the preferred measure of directed forgetting (MacLeod, 1998), the present experiment appeared to have sufficient power, especially with the within-participant design, to detect a directed forgetting effect on false recall if it existed.

**Comparisons to other studies**

Previous studies of free recall using DRM or DRM-type lists have found that the false recall of critical words was generally comparable to the correct recall of studied words (e.g., Read, 1996; Roediger & McDermott, 1995; Toglia et al., 1999). We observed a similar effect, although our recall proportions were somewhat lower than those reported in other studies. However, procedural differences may account for this variation. For example, Read (1996) and Roediger and McDermott (1995) reported correct recall proportions of approximately .65 following a single list presentation, and Toglia et al. (1999) observed a recall proportion of .40 following the study of 5 lists of 12 words. In the present experiment, we obtained recall proportions of approximately .24 for studied words and .26 for critical words, following exposure to 8 or 12 lists of 15 words. The present results are comparable to those observed by Seamon et al. (2001a), who also used 8 lists of 15 words, and comparable in terms of absolute levels of correct recall to those observed by Toglia et al. (1999), as our participants correctly recalled an average of 31 or 36 words following exposure to 120 (8 lists) or 180 (12 lists) words.

**Intrusions of unrelated words**

Aside from the false recall of critical words, participants in this experiment made relatively few intrusion errors in their free recall. Of the total number of words recalled, the average number of noncritical intrusions was low in all conditions. The mean number of intrusions was lower for participants given the remember/remember instructions (8 lists = .88; 12 lists = 1.2) than participants given the forget/remember instructions (8 lists = 1.35; 12 lists = 2.35). But this difference...
was small and due largely to two forget/remember participants in the 12 list condition who produced 10 or more intrusions (excluding these outliers, the mean was 1.6). Overall, the present noncritical intrusion rates for free recall were comparable to those reported in other studies with DRM lists (e.g., Seamon et al., 2001a; Toglia et al., 1999).

**GENERAL DISCUSSION**

In this research, we provided three new observations. First, we showed that the directed forgetting procedure could be applied to the DRM recall task. Second, consistent with prior studies of directed forgetting (see Bjork et al., 1998; Johnson, 1994; MacLeod, 1998), we found that directed forgetting instructions inhibited the accurate recall of studied words that participants were previously instructed to forget, regardless of whether directed forgetting was examined as a within-participant or between-participants variable. Third, and most important, we found that these instructions had no measurable effect on the false recall of critical words, again, whether directed forgetting was examined as a within-participant or between-participants variable. Moreover, the similarity of our results from the within-participant and between-participants designs and the number of lists variable (8 vs 12 lists) demonstrated that these findings were reliable.

In the remainder of this section, we will review our findings in terms of different approaches to false memory, then we will show how these approaches may account for different effects of memory editing on accurate and false memory.

**Different theoretical approaches**

As previously outlined, the implicit activation (Underwood, 1965) and source monitoring (Johnson & Raye, 1981; Johnson et al., 1993) approaches make similar predictions about the effect of directed forgetting instructions on accurate and false memory in the DRM procedure. Specifically, without additional assumptions, both approaches suggest that critical words are treated as studied words and both are remembered in the same manner. False memories, from either perspective, are essentially misattribution errors. If critical words are effectively encoded as studied words, directed forgetting instructions should inhibit both in a similar fashion. The present findings are inconsistent with this interpretation. Directed forgetting instructions inhibited studied words, not critical words. This dissociation suggests that participants in the present research did not treat critical words in the same manner as studied words.

However, our finding of a dissociation is in general agreement with predictions made by fuzzy trace theory (Brainerd & Reyna, 1996, 1998; Brainerd et al., 1999, 2001). As previously noted, this approach holds that studied words and critical words are not treated in the same manner. Studied words are represented by verbatim traces and critical words are represented by gist traces. Given the limited ability of participants to suppress false memory in the DRM procedure (e.g., Gallo et al., 1997; McDermott, 1996; McDermott & Roediger, 1998; Schacter et al., 1998; Seamon et al., 2001b), participants attempting to edit their false memory by recollection rejection, following directed forgetting instructions, should have had limited success. The present findings that directed forgetting instructions inhibited accurate recall, but not false recall, are qualitatively consistent with these predictions. False memories, in fact, survived instructions to forget.

**Memory editing in the DRM procedure**

Before describing how fuzzy trace theory might account for the various effects of memory editing on accurate and false memory in the DRM procedure, we need to review three sets of related findings. First, studies of forewarning have shown that, compared to non-forewarned conditions, forewarning reduced the recognition of critical words and either marginally reduced or had no effect on the recognition of studied words (Gallo et al., 1997, 2001; McDermott & Roediger, 1998). Second, in studies involving multiple study–test trials, accurate memory increased over trials, whereas false memory decreased, whether tested by recall (McDermott, 1966) or recognition (Schacter et al., 1998; Seamon et al., 2001b). Finally, as shown by the present research, directed forgetting instructions inhibited accurate memory, but not false memory. Thus, we can summarise these different results in the following manner: Accurate memory is increased by multiple study–test trials, minimally reduced or unaffected by forewarning instructions, and inhibited by directed forgetting instructions. False memory, on the
other hand, is reduced by forewarning and multiple study–test trials, but unaffected by directed forgetting instructions. Moreover, false memory is not eliminated by any of these memory editing manipulations. To account for these different results, we need to examine fuzzy trace theory’s hypothesised retrieval processes for recognition and recall.

A fuzzy trace interpretation. According to this approach, accurate recognition of studied words is driven largely by verbatim traces, but false recognition of critical words is based predominantly on gist traces. Participants in a recognition memory task access gist and verbatim traces by engaging in familiarity or recollection retrieval processes, analogous to Mandler’s (1980) familiarity or memory search processes (Brainerd et al., 1999, 2001). Participants access gist by first evaluating a test item’s global familiarity to determine if it can be readily accepted or rejected. In addition, they engage in a slower search process of their memory for verbatim traces of specific list items. For studied words, either retrieval process can lead to accurate recognition. But for critical words, the familiarity and recollection processes can lead to different recognition outcomes, depending on the degree of list learning. If list learning is incomplete in the sense that verbatim traces are not available for all list words, participants can falsely recognise a critical word by familiarity because it is highly consistent with the gist of a particular list. Increasing the exposure to the word lists during study, either by repetition (Schacter et al., 1998; Seamon et al., in press) or longer exposure durations (McDermott & Watson, 2001), will lead to more complete verbatim representations for studied words and greater discriminability between verbatim and gist traces at test.

Recently, Brainerd et al. (2001) extended fuzzy trace theory’s dual-retrieval principles to recall. Specifically, these researchers suggested that recall is based on direct access and reconstruction processes. The direct access process most often occurs at the start of recall as participants access the verbatim traces of studied items. When these traces are accessed, participants respond by merely reading this information out of memory. As recall continues, the second process ensues. Recall by reconstruction is slower and less accurate than direct access because the participants attempt to construct the studied items from the meaning or theme of the studied material. Under these conditions, thematically correct gist representations are constructed that can lead to memory errors. Roediger and McDermott’s (1995) finding that false recalls of critical words occurred late in the free recall of DRM lists is consistent with this interpretation.

Fuzzy trace theory’s dual-retrieval principles for recognition and recall can be used to provide a unified theoretical account of the different memory editing effects observed in the DRM procedure. First, research has shown that accurate memory is increased by multiple study–test trials, minimally reduced or unaffected by forewarning instructions, and inhibited by directed forgetting instructions. According to fuzzy trace theory, list repetition should enhance the number of words that have verbatim representations and the completeness of those representations, leading to more accurate recall by direct access or recognition by recollection. The effect on accurate memory of forewarning participants about the false memory effect should be similar to instructions to be cautious about false alarms in general. Because accurate memory is hypothesised to be based on verbatim representations, any reduction in accuracy due to instructions may be attributed to the adoption of a more stringent response criterion for these representations (McDermott & Roediger, 1998). Directed forgetting instructions can inhibit accurate recall if participants suppress the direct access retrieval process that contacts verbatim traces for studied items. Studies of directed forgetting, including the present research, indicate that accurate recall can be inhibited.

Second, research has shown that false memory is reduced by forewarning and multiple study–test trials, but unaffected by directed forgetting instructions. According to fuzzy trace theory, forewarning instructions can reduce false recognition, as the intent of those instructions is to make participants more cognisant of the fact that familiarity-based recognition of gist traces can be incorrect. As a result of the forewarning, participants may become more reluctant to recognise a word simply on the basis of its familiarity. Essentially the same reasoning applies to multiple study–test trials. With list repetition, participants have more opportunity to learn that familiarity-based recognition or reconstruction-based recall can be incorrect. Lastly, directed forgetting instructions may have little or no effect on false recall if the effect of those instructions is to suppress only the direct access of verbatim traces.
Without forewarning or the opportunity to learn that gist traces can lead to incorrect recall in the DRM procedure, reconstruction-based recall may continue to generate thematically correct words when participants are asked to recall all of the previously studied words, even those they were previously told to forget.

Finally, research has shown that false memory is not eliminated by any of the currently studied memory editing manipulations. This observation suggests that participants are unable to limit their recall or recognition to only verbatim representations of the studied items. Gist representations contaminate memory performance by reconstruction processes in recall and familiarity decisions in recognition. The result is substantial false memory even when participants are forewarned, there are multiple study–test trials, or the participants are instructed to forget.

An alternative opponent processes interpretation. Fuzzy trace theory is not the only opponent processes model that can account for the different memory editing results. Recently, Roediger and his colleagues have expanded Underwood’s single process implicit activation hypothesis to an opponent processes model involving activation and monitoring (Balota et al., 1999; McDermott & Watson, 2001; Roediger, Balota, & Watson, 2001; Roediger & McDermott, 2000). Activation is a fast-acting, automatic process that can serve to increase false memory, whereas monitoring is a slower, more strategic control process that can reduce false memory. Underwood’s implicit activation hypothesis is an example of the activation process. Roediger et al. (2001) have suggested that false memory in the DRM task results from the rapid and automatic spread of activation during encoding from representations for list words to semantically associated, but nonstudied, critical words. This fast-acting, automatic process can help explain why participants are able to falsely recognize critical words following extremely rapid list exposures that greatly reduce the recognition of list words (e.g., Seamon et al., 1998). Presumably, very brief exposures to list words are insufficient to activate those representations, but the repeated spread of activation from the related list words to the same semantically associated critical word may summate and yield enough activation to produce a false memory. An automatic activation process may also help explain why comparable levels of false memory are found with young and old adults and adults with mild Alzheimer’s disease in the DRM task (e.g., Balota et al., 1999).

To explain the three different memory editing results, the activation/monitoring approach relies on both processes. First, the increase in accurate memory and the decrease in false memory, over multiple study–test trials, can be accounted for by the increased activation of list word (and critical word) traces over trials, coupled with ever greater learning of list membership, that permits increased monitoring of retrieval by differentiating list words from critical words. Gradually, participants learn which words are in the lists, and which are similar to list members, but not present (Benjamin, 2001; McDermott & Watson, 2001; Seamon et al., in press). Second, the reduction in false memory due to forewarning instructions can be attributed to the increased use of monitoring, as participants strategically try to determine the critical words during study so as to avoid reporting them at test (Gallo et al., 1997). Any minor effect of forewarning instructions on accurate memory may be attributed to the use of a more conservative response criterion. Third, the reduction in accurate memory, but not false memory, due to directed forgetting instructions can be attributed to differences in trace strength activation for list words and critical words. Critical word representations may receive greater activation than list word representations due to the repeated spread of activation following exposure to the semantically related list words (Seamon et al., 1998). Consistent with this view, research has shown that false memory for critical words increases with increases in the number of semantically related list words at study (Robinson & Roediger, 1997), and false recall for critical words has demonstrated greater persistence than accurate recall for list words over different retention intervals, (Seamon et al., 2001a; Thapar & McDermott, 2001; Toglia et al., 1999). Just as critical words are more resistant to forgetting over retention intervals, they are also more resistant to forgetting by directed forgetting instructions. Thus, like fuzzy trace theory, the activation/monitoring approach can account for the different memory editing results, including those of the present research, in terms of opponent processes that serve to increase and decrease false memory. At present, however, fuzzy trace theory is a more thoroughly articulated description of memory processes than the activation/monitoring approach, it generates quantitative, as well as qualitative predictions, and it has been applied to a variety of memory tasks (e.g., Brai-
nerd & Reyna, 1998; Brainerd et al., 2001; Brainerd, Wright, Reyna, & Payne, in press).

**Final thoughts**

Why are false memories so difficult to inhibit? Several factors may contribute to their robustness. First, in most studies involving the DRM procedure, multiple lists are tested after a single study trial. Because list learning is not perfect, editing of false memory is bound to be limited. Second, even when multiple lists are studied frequently (e.g., 25 times by Seamon et al., in press), false memory can still occur, as numerous lists yield numerous critical words that must be maintained and segregated in order to be rejected at test. Finally, Brainerd et al. (2001) suggested that editing of false memory could occur automatically, outside of a participant’s conscious awareness, or strategically, within that person’s awareness. Our finding that directed forgetting instructions inhibited accurate recall, but not false recall, indicates that our participants tried to follow those instructions. But whatever strategic process they employed with the studied words was notably unsuccessful with the critical words. This finding is reminiscent of previous research that showed that, following extremely rapid list exposures, participants falsely recognised critical words even when they were unable to recognise studied words (Seamon et al., 1998, 2000). To date, the best explanations for such memory errors are found in opponent processes accounts such as fuzzy trace theory and the activation/monitoring approach. False memory may be an inherent product of a memory system that allows for memory editing, but does not provide complete monitoring and control of its associated retrieval processes. Thus, in many instances, participants may not be able to report the whole truth and nothing but the truth.

Manuscript received 8 January 2001
Manuscript accepted 29 August 2001

**REFERENCES**


