Assessing activation of true and false memory traces:

a study using the DRM Paradigm

Ida Sergi, Vincenzo Paolo Senese, Maria Pisani, & Giovanna Nigro

Department of Psychology Second University of Naples

Author Note

Corresponding author
Ida Sergi, Ph. D.
Department of Psychology
Second University of Naples
Viale Ellittico, 31, 81100 Caserta
e-mail: ida.sergi@unina2.it

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
Abstract

This paper compares the activation of false memory traces with true memory traces and control items. An incidental learning version of the Deese-Roediger-McDermott paradigm (DRM) is used in conjunction with a Lexical Decision Task (LDT). One hundred participants read 5 DRM lists intermixed with 4 non-word lists. After a delay manipulated into 3 experimental conditions (immediate, 3-minute and 10-minute), they were asked to classify actually studied items, CIs, new words and non words in an LDT. Results showed that, independent of the delay, the classification latency of the CIs was the same as that for actually studied words and that both were more active than matched control items. Moreover, results showed that the delay affects the activation of true and false memories at the same rate. Overall the results seem to support the hypothesis that false memories trace becomes additive traces that are integrated into the episodic memory, with the same features of the true memories.

Keywords: False memory; Activation level; Lexical decision task; DRM
Introduction

False memories are defined as the remembering of events that never happened or as remembering them quite differently from the way they happened (Roediger & McDermott, 1995).

In the literature, one of the most frequently used paradigms to study false memories is the Deese-Roediger-McDermott paradigm (DRM; Deese, 1959; Roediger & McDermott, 1995). In a typical DRM experiment, participants study a list of words (e.g., hot, snow, winter, etc.) that are all related to a non-presented word (the critical item [CI], e.g., cold). High levels of false memories for the CI have been demonstrated when participants are asked to recall or recognize the words of the lists (Gallo 2006).

In the literature the most used theoretical explanation of the DRM memory illusion is the Activation/Monitoring Theory (AMT; Roediger, Balota, & Watson, 2001). The AMT suggests that false memories are due to a combination of spreading activation and a more controlled monitoring process (Roediger et al., 2001). During encoding, the study of words associated to the CI indirectly activates the CI representation, which becomes integrated into the episodic memory trace and is associated with the encoding context as well as the studied items (Hicks & Hancock, 2002). During subsequent testing, subjects produce false memories because they fail to distinguish between items generated internally and those presented to them externally. Then, according to Roediger and colleagues, false memories are produced by the conjunction of indirect activation of non studied CI and failure of source monitoring (Roediger et al., 2001).

The second major framework used to account for false memory in the DRM paradigm is the Fuzzy Trace Theory (FTT; Brainerd & Reyna, 2002, 2004). The FTT assumes that participants encode information by a two-track process, namely, verbatim and gist. These traces are qualitatively different from each other: verbatim traces represent the surface form of the experienced item, whereas the gist trace represents its semantic content, including meaning, relations and patterns (Brainerd & Reyna 2002). Therefore memory performances are based on the retrieval of both
verbatim and gist traces. In this view, both traces should support true memories, while should have opposite effects on false memories. According to this model, false memory observed into DRM experiments is attributed to gist extraction that occurs during encoding, while the absence of the verbatim traces should suppress or reduce the probability of false memories (Brainerd, Wright, Reyna & Mojardin, 2001).

These two approaches share the idea that a memory representation corresponding to CI is stored during encoding. The difference between the two theoretical approaches refers to the conceptualization of the underlying memory traces. The AMT assumes that the representation of CI contains both semantic and surface information, just like the representations of list items, and that the CI trace is integrated into the episodic memory experience with the same features of the studied items (Reder, Donavos, & Erickson, 2002). In contrast, the FTT assumes that representations of list items include both semantic and surface information, whereas the gist trace corresponding to the CI includes only semantic information.

Given the theoretical relevance of investigating the differences between true and false memory traces, the aim of this paper was to examine and compare the status of the underlying memory traces of true and false memories by means of their activation level.

In the literature, to compare the activation level of true and false memories, different implicit memory paradigms have been used (e.g., word fragment completion, lexical decision task, McDermott, 1997; McKone & Murphy, 2000; Tse & Neely 2005). In this study we focused on studies that used the lexical decision task (LDT) because it can be considered a relatively pure measure of how much activation the study of a DRM list produces for studied items and CIs (Tse & Neely 2005). Moreover, it minimizes any contamination from explicit memory and source monitoring (Meade, Watson, Balota, & Roediger, 2007). In the LDT, participants are presented with strings of letters and are asked to classify them as words or non-words. The underlying logic is
that there is an inverse relationship between LDT latency and the corresponding trace activation (Anderson 1983).

Thus far, studies that have used the LDT to measure the activation of DRM-item memory traces have produced controversial results (Tse & Neely, 2005). A possible explanation of the observed discrepancy is that the LDT parading has been implemented in different ways. For example, in some studies the LDT was administered after the presentation of all the DRM lists while in other studies the LDT was administered after the presentation of each DRM list. In some studies the LDT presented pseudohomophone non-words while in the other studies the LDT did not present pseudohomophone. Below we present a brief overview of the studies.

Zeelenberg and Pecher (2002, Experiment 3) measured the activation of CIs by adapting McDermott’s (1997) procedures. They presented 18 classical DRM lists studied incidentally or voluntarily. The LDT was administered at the end of the lists’ presentation. Independent of the encoding condition, they found that CIs were less active than studied items. Using a similar paradigm, McKone (2004) asked participants to study voluntarily 16 DRM lists and administered a final LDT after 3 or 10 minutes. Also in this study, independent of the experimental condition CIs were less active than studied items. In both Zeelenberg and Pecher (2002) and McKone (2004) studies the DRM lists were presented one item at a time at a rate of 2 s or less, the LDT was administered at the end of DRM lists presentation, and in the LDT non-words were not pseudohomophone.

In contrast with these results, by adopting a similar procedure, Whittlesea (2002) observed that CIs and list items were equally active. In his study participants were asked to study voluntarily 18 DRM lists then a final LDT was administered. Differently from Zeelenberg and Pecher (2002) and McKone (2004) studies, Whittlesea presented simultaneously all the items of each DRM list, and the LDT compounded pseudohomophone non-words. Similar results were also observed by Senese, Sergi, and Iachini (2010). In this latter study, authors utilized an incidental encoding task
for a single DRM list. They asked to participants to read the DRM items and no mention was made
to any memory task. A single classical DRM list was presented followed by a final LDT. Results
showed that the CI trace was equally active as traces associated with matched studied items and
more active than matched new-item traces. In this study, the single items of the DRM list were
presented one at time at a rate of 2 s, and in the LDT non-words were pseudohomophone.

Studies that administered the LDT after the presentation of each DRM used pronounceable
non-words in the LDT and showed that false items were more or equally active that studied items.
The only exception was the study of Meade et al. (2007) that presented 48 modified DRM lists. In
their study, authors included non-words in the original DRM lists to reduce the utility of retrieving
the context of earlier lists in the execution of the LDT. Results showed that the classifications of CIs
were facilitated only when tested about one second after the studied list.

Hancock, Hicks, Marsh, and Ritschel (2003) presented 25 DRM lists manipulating the
number of items in each list and the relative backward associative strength indices. The LDT was
given after a 30 s math filler task that followed each DRM list presentation. Results showed that,
independent of the experimental condition, false memory traces were more or equally active than
matched studied items. In this study the DRM lists were presented one item at a time at a rate of 5 s,
and in the LDT non-words were pronounceable non-words created by changing a consonant or a
vowel. Similar results were founded by Tse and Neely (2005, 2007). In two different studies they
presented 25 classic DRM lists with intentional learning instructions. For each list, the LDT was
administered after a delay of 30 sec and the non-words of the LDT were pseudohomophonic. In
both studies, results showed a facilitation effect for CIs.

To explain their results Meade et al. (2007, Meade, Hutchison & Rand, 2010) considered
that it was possible that the delay between the study of the DRM lists and the LDT administration
can affect the activation of the false trace. In their view, the activation of CIs should decay faster
than the activation of studied items. In line with this consideration, when the activation level of CIs
is evaluated immediately after the study of the DRM lists, CIs should be as active as studied words. Conversely, if the activation of CIs is tested after some amount of time or after the study of more than one DRM list, the CIs should be less active than the true items.

In summary, comparing the above results, it seems that the facilitation effect of the CIs is observed when the LDT is administered after each DRM list presentation, but only with classical DRM lists, or when the LDT is administered at the end of the lists’ presentation, and only if pseudohomophone non-words are used. In our view studies that administered the LDT after each DRM list presentation could have a bias of a possible contamination from explicit memory and source monitoring (Meade, Watson, Balota, & Roediger, 2007). Indeed because participants were administered a cycle of study-test phases they could understand the manipulation of the experiment, than recognizing the relation between the two phases. Thus we think that it is important to disentangle if the same facilitation effect of the CIs is observed also if the subject is completely unaware of the experimental manipulation so reducing any contamination of the LDT.

The aim of this study is to compare the activation levels of traces associated with CIs and true items by means of an LDT compounded of pseudohomophone non-words and administered at the end of the lists presentation. Moreover, we investigate whether the activations of both types of items decay at the same rate. To compare the status of false and true memories traces, an LDT was administered after the presentation of nine lists (five classical DRM lists and four non-word lists) and before the final explicit recognition task. We included the non-word lists to minimize the strategy of using the presence of an item in the study list as a cue in the LDT. In the LDT, the participants were presented with different types of items: the actually studied words, the CIs, the actually studied non-words, the matched new words and the new non-words. Both kinds of non-words were pronounceable. To investigate the hypothesis that the activation of true and false memories decay at different rate, and to compare our results with McKone (2004), we manipulated the delay between the study and the test phases in three experimental conditions. In the first
condition, the participants performed the LDT immediately after the study of the lists. In the second condition, the LDT was performed 3 min after lists study. In the last condition, the LDT was performed 10 min after lists study. Finally, to compare the activation levels of item lists, CIs and matched words in the absence of any differences among them, a control condition in which the participants uniquely executed the LDT was considered.

In accordance with the literature, we assumed that the time necessary to classify words in the LDT is an inverse function of the level of activation of the underlying representation (Anderson, 1983).

In line with the AMT, if it is true that the activation process involved in the DRM create a false memory trace of critical items, one should expect that after the study phase and over longer delays CIs are as activated as studied items. On the contrary, in the view of FTT, a different pattern is expected, if the activation process creates uniquely a short lasting activation of CIs (Meade et al., 2007, 2010), one should expect that immediately after the study phase CIs traces are as active as studied items, whereas over longer delays CIs traces are less active than studied items. Based on the results of previous studies that showed a similarity between CIs and studied items, we hypothesized that in all experimental conditions the classification latencies of actually studied words would be similar to the CIs and shorter than matched new words and non-words.

Method

Participants

A total of 100 undergraduate university students (54 females and 46 males) participated in the experiment. Their ages ranged from 18 to 35 years ($M = 21.5$ years, $SD = 3.5$). All participants were native Italian speakers with normal or correct-to-normal vision. They were matched by gender and age, and randomly assigned to one of the four experimental conditions and tested individually in sessions that lasted approximately 25 minutes.

Materials
In the study phase, nine lists were used: 5 DRM lists and 4 non-word lists. The 5 DRM lists were developed by a preliminary study according to the DRM paradigm (Stadler, Roediger, & McDermott, 1999) in the Italian population. The pilot study has indicated that they have a high BAS index and therefore high probability of inducing false memories (mean BAS index of the lists was .35). Each list contained 15 semantic words associated with a related critical item. Each DRM list was followed by a list of pronounceable non-words. In the LDT, participants were presented with letter strings and were asked to classify them as correctly spelled Italian words or as non-words. There were 80 letter strings to be classified: 15 items from the DRM lists (3 from each DRM list: from the I, VIII, and XII position), the 5 CIs, 15 non-words presented during the study phase (3 from each list), 25 new non-words and 20 new words. The non-words were Italian orthographically regular words and created by changing a consonant or a vowel in Italian words. The new-words were selected according to the following criteria: they were Italian words semantically unrelated to the items of the studied list or to the CIs, and, on the basis of normative data for words in Italian (CoLFIS; http://www.istc.cnr.it/material/database/colfis), they were matched for length, syllable number, and word frequency to the lists items and to the CIs. We gave particular care to the matching procedure because length, syllable number, and word frequency are known to affect latency independently of activation status (Hancock et al., 2003). To avoid a possible semantic additive effect due to the re-presentation of the lists items in the LDT, the CIs were presented before each studied item.

Procedure

The experiment consisted of two phases. In the first phase, participants were asked to read the items of the lists, which were presented one at a time at the centre of the computer screen for 2s each. The nine lists were presented as one long list of 135 items. No mention was made of a memory test. In the second phase, the LDT instructions were shown on the screen. Participants were presented with letter strings and were asked to classify as quickly and as accurately as possible
whether the letter strings formed a correctly spelled Italian word or a non-word. The presentation of items during the LDT was random with two exceptions: none of the items of interest appeared among the first four items, and all the CIs were presented before all the items of the studied lists. According to the experimental conditions, participants in the immediate condition performed the LDT immediately after the study phase. In the 3 min and 10 min conditions, participants were administered a filler math task before the LDT. In the control condition participants performed only the filler math task before the LDT.

Results

Incorrect lexical decision task responses were removed from the analyses, and RTs faster than 200ms and slower than 2000ms were excluded. Univariate distributions of the observed RTs for mean scores of each type of item were examined for normality (Shapiro & Wilk, 1965). The results indicated that univariate normality did not hold. A logarithmic transformation was used to normalize the distributions (Tabanick & Fidell, 1996). Performances on the LDT were analyzed by conducting a 5×4 mixed factorial ANOVA that treated Word Type (Words lists, Non-words List, CIs, matched Non-words, and matched Words) as within factor and Condition (immediate, 3 min, 10 min and control) as between factor. The analysis was performed on the transformed variables, but for descriptive purposes, untransformed data were used to report means. The Bonferroni correction was used to analyze post hoc effects, and the magnitude of significant effects was indicated by partial eta squared ($\eta^2_p$). To test if effects were influenced by the gender, we replicated the analyses using the gender as between-subjects factor. Results did not change.

The 5×4 factorial mixed ANOVA performed on the LDT data showed that the latency of the LDT was influenced by the Word Type, $F(4,384) = 242.522; p < .001; \eta^2_p = .716$, by the Condition, $F(3,96) = 3.444; p = .020, \eta^2_p = .097$, and more germane here by the interaction Condition×Word Type, $F(12,384) = 2.376; p = .006, \eta^2_p = .069$. The post hoc analyses for the Word Type effect revealed that there were no significant differences between the mean latency for classifying actually
studied items ($M = 682.5$) and the CIs ($M = 684.3$), and both means were shorter than the mean latency for classifying the matched new words ($M = 736.3; ps < .001$), the studied non-words ($M = 1024.5$) and non-words ($M = 957.5; ps < .001$; Table 1). The post hoc analyses for the Condition effect revealed that the mean latency for classifying items in the 10 min condition ($M = 893.0$) was greater than both immediate and control conditions (respectively $M = 779.8$ and $M = 803.1, ps < .05$). The post hoc analyses for the Condition×Word Type effect revealed that in the immediate, 3 min and 10 min conditions the mean latency for classifying actually studied items and the CIs did not differ significantly and that both the mean latencies were shorter than the mean latency for classifying the matched new words, the studied non-words and non-words ($ps < .01$; Table 1). In contrast, in the control condition, the data showed no significant differences among the mean latencies for classifying words list, the CIs and the matched new words. They were all shorter than the classification latency of both the types of non-words ($ps < .001$).

(Table 1 about here)

Discussion

The aim of this study was to compare the activation level of memory traces associated with false and true memories in the DRM paradigm and to examine whether the activation of CIs and true items follows the same pattern of decay. To compare the traces, the DRM paradigm was used in conjunction with the LDT paradigm.

The results of the LDT showed that, in general, the CIs were characterized by classification latencies equal to those of actually studied words and that both types of items were classified with shorter latencies than matched new words and non-words. The interaction effect showed that this pattern was identical for the three experimental conditions and confirmed that in the control condition there weren’t any latency differences among the three types of words. These results suggest that after the encoding phase and at least until a 10 min interval, false memory traces have the same activation status as true memory traces and are more active than any other matched non-
studied items. The results relative to the control condition confirmed that, in the absence of a study phase, there are no differences in the classification latency between the list items and the matched words.

In the literature, previous researches that used the LDT to measure activation of memory traces of studied DRM items and CIs produced controversial results (Tse & Neely, 2005). Recently, Meade et al. (2007, 2010) suggested that a possible explanation of the controversial results could be the delay between the study of DRM lists and the LDT. According to them, the activation of CIs should decay faster than that of the real items. In this paper, we tested this hypothesis directly by manipulating the delay between the study of DRM lists and the evaluation of activation traces. Our results do not support the Meade et al. (2010) hypothesis and showed that, if we compare the activation status of true and false memories, the underlying memory traces are equally activated and follow the same decline over the time. In addition, our results do not support Zeelenberg and Pecher (2002) and McKone’s (2004) results because they found no facilitation for the non-studied CIs. This latter discrepancy may be due to a procedural factor. Indeed, in these studies the DRM lists were not developed for the specific population as the DRM paradigm prescribes (Stadler, Roediger, & McDermott, 1999) but were adapted from Roediger and McDermott (1995), and this could have reduced the activation of the associated CIs. Moreover in both studies no pseudohomophone non-words were used in the LDT. On the other hand, our results are in line with the studies that found a clear facilitation in the LDT of the non-studied CIs (Hancock et al., 2003; Senese et al., 2010; Tse & Neely, 2005, 2007; Whittlesea, 2002) and indicate that the activation status of false memory traces is not as short lived as suggested by Meade et al. (2007). As argued by the same authors, a possible explanation of the discrepancy is that the facilitation effect of CIs lasts longer when the blocked lists are used, as in our experiment, than the mixed lists containing both non-words and unrelated words. Moreover, as suggested by Tse & Neely (2005, 2007), because DRM lists include multiple, semantically related words converging on a single CI, it is possible that the activation
produced by classical DRM lists may last longer than the activation produced by single associates.

In our study we maximized this effect by choosing DRM lists that elicit high false recognition rates for CIs in explicit memory tests, and by using pseudohomophone non-words. To our knowledge, this is the first study that showed that the similarity between the activation status of true and false memory traces lasts for a 10 min interval, in line with false memory studies that showed high levels of false recognition on delayed explicit memory tests (see Gallo, 2006, for a review).

From a theoretical perspective these results have relevant implications. In fact, as mentioned in the introduction, two models have been proposed to explain the false memory effect with the DRM paradigm: the Activation/Monitoring Theory and the Fuzzy Trace Theory. Both the AMT and the FTT assume that true and false memories are phenomenologically similar but can be differentiated by the conceptualization of the underlying memory traces. That is, in the FTT the underlying representations of true and false memories should be different, while in the AMT the false memory trace should be equivalent. In this perspective, our results on activation level of memory traces support the idea that the true and false memories traces are equivalent, as hypothesized by the AMT. On the contrary, if false memories consisted purely of gist traces, as suggested by the FTT, it would be unlikely that such effects would consistently emerge.

In conclusion we can affirm that the overall results seem to support the hypothesis the false memories trace becomes an additive trace that is integrated into the episodic memory, with the same features of true memories (Reder, Donavos, & Erickson, 2002; Roediger, McDermott, Pisoni & Gallo, 2004). This effect can be accounted by the combination of the associative processes active during the encoding phase and by the effect of the encoding specificity principle (Tulving & Thomson 1973). In this sense, in line with the AMT, we believe that the activation of the CI during the episodic experience makes it a good candidate for a false recall or recognition. The current results therefore echo the ideas initially proposed by Jacoby and colleagues in raising the intriguing
possibility that processes occurring outside our conscious awareness can be an important
determinant of false memory creation (Cotel, Gallo & Seamon, 2007).

References


Psychological Science*, 11, 164-168


phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and
Cognition*, 27, 341-361.

sufficient to produce false memories. *Consciousness and Cognition*, 17, 210-218.

*Journal of Experimental Psychology*, 58, 17-22.


Hancock, T. W., Hicks, J. L., Marsh, R. L., & Ritschel, L. (2003). Measuring the activation level of
critical lures in the Deese-Roediger-McDermott paradigm. *American Journal of Psychology*,
116, 1-14.

Hicks, J. L., & Hancock, T. W. (2002). Backward associative strength determines source


Table 1

Mean latency (in Milliseconds) in the LDT trials as a function of Word Type and Condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Words List</th>
<th>List</th>
<th>CLs</th>
<th>Matched new Words</th>
<th>New Non-words</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate</td>
<td>656.3(^a)</td>
<td>956.9(^b)</td>
<td>648.9(^a)</td>
<td>703.8(^c)</td>
<td>932.9(^b)</td>
<td>779.8(^1)</td>
</tr>
<tr>
<td>3min</td>
<td>682.2(^a)</td>
<td>989.4(^b)</td>
<td>699.9(^a)</td>
<td>761.0(^c)</td>
<td>972.3(^b)</td>
<td>820.9(^1,2)</td>
</tr>
<tr>
<td>10min</td>
<td>712.8(^a)</td>
<td>1154.2(^b)</td>
<td>712.7(^a)</td>
<td>782.7(^c)</td>
<td>1102.7(^b)</td>
<td>893.0(^2)</td>
</tr>
<tr>
<td>Control</td>
<td>678.7(^a)</td>
<td>997.6(^b)</td>
<td>675.9(^a)</td>
<td>697.6(^a)</td>
<td>965.5(^b)</td>
<td>803.1(^1)</td>
</tr>
<tr>
<td>Mean</td>
<td>682.5(^a)</td>
<td>1024.5(^b)</td>
<td>684.3(^a)</td>
<td>736.3(^c)</td>
<td>993.3(^b)</td>
<td>–</td>
</tr>
</tbody>
</table>

Note: \(^a\)Equal letters or equal numbers indicate equal means (\(p > .05\)) in the post hoc analysis with the Bonferroni correction.