A memory-interference versus the “dud”-effect account of a DRM false memory result: Fewer related targets at test, higher critical-lure false recognition

Jerwen Jou1 · Mark Hwang2

Accepted: 3 March 2022 / Published online: 22 March 2022
© The Psychonomic Society, Inc. 2022

Abstract
Memory interference theories hold that exposure to more similar information to a target item impairs memory of the target item. The dud effect refers to the finding in eyewitness lineup identification that fillers dissimilar to the suspect cause more false identification of the suspect than similar fillers, contrary to the interference concept. Previous studies on the Deese–Roe-diger–McDermott false memory typically showed a testing priming effect that a larger number of studied items presented at test leads to a higher level of false recognition of the critical lure (CL). In the present study, either all, or all but one studied item were replaced by unrelated distractors at test. Subjects made more false recognitions of the CL in the no- or only-one-studied item than in the multiple-studied-item condition, supporting the dud-effect account. The slower response time in the “dud” condition suggested a deliberate, monitoring-like approach taken by subjects in that condition.

Keywords Dud effects and false memory · Dud effects in DRM paradigm · DRM test list composition effect

It has been well established in memory literature that items on the memory list or in the test that are semantically (Gil-lund & Shiffrin’s, 1984, SAM; Hintzman’s, 1988, MIN-ERVA; Shiffrin & Steyvers’s, 1997, RAM) or otherwise similar to the target items (e.g., Conrad, 1964 for sound similarity) cause more interference for memory of target items than dissimilar items. For example, in the well-known release from proactive interference paradigm (Watkins & Watkins, 1975; Wickens et al., 1963; Wickens, 1972), recall accuracy decreases progressively with continued learning and testing of additional items from the same taxonomical category, a phenomenon known as proactive interference. But when the learning-testing materials switch to a different category at the last trial, memory rebounds to the initial high level. Other memory theories such as item-noise (e.g., Shiffrin & Steyvers, 1997), context-noise (e.g., Dennis & Humphreys, 2001), or item-context noise models (Criss et al., 2011) propose a similar concept in that items on the study list or test list or both study and test lists that are similar to the target probes cause more memory interference and consequently more performance decrement for the targets than do dissimilar items. Although these models make different claims as to the specific sources of the interference, they agree that the more similar the items are to the target probes, the more noise, or memory interference, the target probes suffer.

Interestingly, however, in a different but related field of study, the eyewitness lineup identification, the findings appeared to show a contrary pattern. When the fillers (the nonsuspects put on the lineup to test the eyewitness’ ability to discriminate the suspect from these fillers) have characteristics that are highly dissimilar to the suspect and hence are highly implausible candidates for selection (thus called duds), the rate of selecting the suspect and the subjects’ confidence for the selected suspect relative to the non-dud condition where the fillers are similar to the suspect (Charman et al., 2011; Hanczakowski et al., 2014; Windschitl & Chambers, 2004). There are a dud-addition and dud-replacement conditions both of which increase the selection rate and confidence for the selected suspect relative to the non-dud condition. The dud-addition condition involves adding several duds to a shorter non-dud lineup.
The dud-replacement condition involves replacing several non-dud fillers with duds, thus keeping the lengths of the two lineups unchanged (Charman et al., 2011). There are different specific hypotheses on the cause of the dud effect (Charman et al., 2011; Windschitl & Chambers, 2004). In general, the effect can be characterized as a contrast effect. That is, as the distractors become highly implausible, the focal item, or the suspect 1, becomes relatively more plausible and hence gains in subjective (or false) likelihood of being the target.

We adopted a modified Deese-Roediger-McDermott (DRM; Roediger & McDermott, 1995) paradigm to test the interference theories against the dud-effect account of a finding from a task in that paradigm. Roediger and McDermott revived a memory research paradigm first developed by Deese (1959) in which people study lists of semantically associated words, such as night, tired, dream, blanket, snore, bed, slumber, etc. with their meanings all converting on the theme of sleep, without the theme word itself (i.e., the critical lure or CL henceforth) presented on the study list. At the test, subjects falsely recall or recognize the CL at a level comparable to that of a studied word located at the middle position of the studied list (Roediger & McDermott, 1995). The very high and robust false recall and recognition of the CL has since been known as false memory or memory illusion (Roediger, 1996), and the method referred to as the DRM paradigm (Gallo, 2010). In a standard DRM task, there are 15 semantic associates on a study list, and six test probes on the corresponding test list. Of the six probes, three are studied words (e.g., bed, night, tired) and the other three probes are the CL (sleep) and two unrelated distractors (e.g., banana, shoe). Thus, the three target probes are related to the list theme and the CL, and the two distractors are unrelated to the list words or the CL. In the design of the present study, the control condition adopted the standard DRM test list makeup. In the experimental condition, on half of the lists, the three target words were replaced by three unrelated distractors (equivalent to duds), thus leaving no target words on those test lists, and the CL as the only related probe. On the other half of the lists, only two of the three target words were replaced by unrelated distractors, thus leaving one target word, one CL, and four unrelated distractors on a test list. Although there is a difference between our and the dud-replacement manipulation, we think that the underlying logic of our manipulation is essentially the same as the dud-replacement manipulation. In the dud-replacement manipulation, plausible fillers are replaced by duds.

In our manipulation, target probes (hence, highly plausible choice candidates) were replaced by unrelated distractors (equivalent to duds). We believe that there is an essential analogy between the two manipulations despite the specific difference.

According to the interference theories, the standard test-list design in the control condition with three target words should generate more memory interference than that in the experimental condition with only one related probe, the CL, on half of the lists, and one CL and one target on the other half of the lists. This idea is consistent with the spreading-activation theories’ account which assumes that exposure to more related items will lead to more implicitly activated responses (Collins & Loftus, 1975; Roediger et al., 2001a, b; Underwood, 1965). However, the dud account makes the opposite prediction in this case, i.e., the unrelated distractors that replace the target probes on the test list will enhance the plausibility of the CL and lead to an increase in the false acceptance of the CL.

The findings regarding the effect of increasing the number of studied words at test on the rate of FA to the CL (referred to as the test-induced priming) were inconsistent. Marsh et al. (2004) who presented the CL at test after zero, three, or six list items were presented found no effect of this manipulation on the FA rate of the CL. Likewise, Dodd et al. (2006), using an auditory modality and presenting 0 to 5 studied items before the CL at test, found no effect on the FA rate to the CL. On the other hand, Coane and McBride (2006) found that increasing the number of studied words from zero to six at test before the presentation of the CL increased the FA rate of the CL, although further increasing them beyond six made no difference. Dewhurst et al. (2011) found an effect of increasing the number of studied words when a monitoring mechanism was not invoked. As well, Diez et al. (2004) by increasing the number of studied items from zero to four before the presentation of the CL found an effect under a speeded instruction, but no effect under a nonspeeded condition. Similarly, Marsh and Dolan (2007) found such an effect in RT in a self-paced test, but one in an increased CL FA rate when the recognition decision was made under a speeded instruction. These three findings support the idea that additional spreading activation effect arises from related test probes when the monitoring process is suppressed. Inconsistent to the finding of an increased FA rate to the CL due to the presence of more studied words at test, Jou et al. (2018) found that when the two unrelated distractors on the test list of the standard DRM paradigm were replaced by related ones, the FA rate to the CL actually decreased. This effect can be considered to be a reverse-test-induced effect. That is, more related distractors at test actually reduced FAs to the CL. What will happen to the FA to the CL if the two of

---

1 There are two types of suspects, an innocent and a guilty suspect. Here we are talking about an innocent suspect who looks similar to the perpetrator.
the three or all three target probes are replaced by unrelated distractors? As noted above, according to the previous findings, fewer studied words presented at test lead to a lower rate of FA to the CL, which is also the prediction of the memory interference theories. However, the dud account makes a contrary prediction that fewer studied probes (or more unrelated distractors) at test will increase the CL FA rate.

Method

Subjects

Sixty introductory psychology students participated in the control condition and 132 students participated in the experimental condition for fulfilling a partial course requirement at the University of Texas–Rio Grande Valley. Seventy-two percent of the subjects was female. Their mean age was 20.78 years, ranging from 18 to 45.

Design, materials, and procedure

The 24 lists of semantic associates used in Roediger and McDermott’s (1995) study were adopted in this study. Each list had 15 semantic associates all related to a theme. The 24 lists were divided into three sections, with the first eight lists being assigned to Section 1, the second eight lists to Section 2, and the third eight lists to Section 3. One third of subjects studied and were tested on Section 1, one third on Section 2, and one third on Section 3.

The study lists were the same across the control and the experimental conditions. A test list comprising six probes was constructed for each list of the semantic associates. The compositions of the test lists were different between these two conditions. In the control condition, the basic structure of the test list of the DRM paradigm was followed—that is, three of the six test probes were target (studied) words, so the target presence rate was 50%. In the experimental condition, half of the eight lists had test lists in which no target was present. Only the CL was the probe related to the semantic theme of the list concerned. Hence, for these lists, the target presence rate was 0%. The other four lists each had one target word, one CL, and four unrelated distractors. Therefore, for these lists, the target presence rate was 1/6 or 16.66%. Thus, overall, for the experimental condition as a whole, the target presence rate was 8.33%. The one-target and zero-target lists were counterbalanced across subjects. The goal of setting up the target presence in this way in the experimental condition was to minimize the target presence rate without giving the subjects the impression that none of the test lists had a studied word. The target words were randomly selected from the respective studied associate lists. The unrelated distractors were randomly selected from the 16 unstudied lists, not excluding the CLs of these lists with the condition that these distractors for a test list be semantically unrelated to each other and to the theme of the studied list. During the test, the six test probes were presented one at a time in a new random order for each subject and for each list both in the experimental and in the control condition. Although the position in which the CL occurred at test was not controlled, on average, it should be preceded by more unrelated probes in the experimental than in the control condition.

During the study phase, the 15 list words were presented one at a time in a random order, for 2.5 s per word, with a 1-s blank screen between the words. At the end of the presentation of a list, subjects did backward counting from a three-digit number for 20 s. They were instructed to count at a reasonably fast speed. When they made an error, the program made a beep and issued a warning urging them to be more careful. During the test phase, the 6 test probes were presented one at a time in a random order for a recognition. Subjects pressed “z” key for a “Yes” response for a word they judged to have been studied, and the “f” key for a word they judged to be new. The mapping of the response keys to the yes/no responses was counterbalanced across subjects. Thus, for each list, they studied 15 words, counted backward for 20 s, and took the recognition test. They repeated that process eight times.

Subjects downloaded the Python program from a website to run the experiment on their own computers. Before the start of the experiment, subjects were told to be as fast and as accurate as possible in their responses. They were told to
find a quiet room to do the experiment, not to pause or rest during the test, and to avoid any distractions. They could take a short break at the end of a test before beginning the study for the next list.

Results

The first analysis used the first 60 data sets received from the experimental condition. The analysis of variance (ANOVA) indicated that the mean FA rate to the CL in the experimental condition (.708) was significantly higher than that in the control condition (.562), $F(1, 118) = 8.24$, $p = .005$, $\eta_p^2 = .065$. The target presence versus absence factor within the experimental condition did not have a reliable effect on the FA rate for the CL ($\eta_p = .704$; target absence = .717), $F(1, 59) = .09$, $p = .763$. The hit rate to the target words in the experimental condition (.816) was a little lower than that in the control condition (.838), but the difference was not significant, $F(1, 118) = .50$, $p = .482$. The correct rejection rates to the unrelated distractors of the control and the experimental condition were very close (.965 for the control and .967 for the experimental condition) and the difference was nonsignificant, $F(1, 118) = .01$, $p = .937$.

Analyses of RT data

RTs equal to 0 or to 15,000 ms or longer were considered errors and excluded from analysis. These trials made up of 1.21% of the total data. A condition (control vs. experimental) by probe type (CL vs. target vs. distractor) ANOVA indicated that the experimental condition overall mean RT (2,099 ms) was significantly higher than the control condition overall mean RT (1,583 ms), $F(1, 118) = 7.52$, $p = .007$, $\eta_p^2 = .060$. The main effect of probe type was significant, $F(2, 236) = 38.81$, $p < .0001$, $\eta_p^2 = .247$. A Newman–Keuls post hoc comparison showed that the three RT means for the three probe types (CL = 2,217 ms; target = 1,795 ms; distractor = 1,511 ms) were significantly different from each other. The condition by probe type interaction was marginally significant, $F(2, 236) = 2.74$, $p = .067$, $\eta_p^2 = .023$. As shown in Fig. 1, the interaction derived from the RT differences between the two conditions being larger for the CL and the target than for the distractors.

When only the FA RTs to the CL were compared between the control and experimental conditions, the mean RT of the experimental condition (1997 ms) was significantly longer than that of the control condition (1,471 ms), $F(1, 113) = 9.19$, $p = .003$, $\eta_p^2 = .075$. When the correct rejection RT of the CL was compared across the control and the experimental conditions, the latter (3,407 ms) was also significantly longer than the former (2,536 ms), $F(1, 97) = 5.61$, $p = .019$, $\eta_p^2 = .055$. A comparison of the FA RT to the CL between the target presence and absence conditions within the experimental condition showed that the FA mean RT of target presence condition (1,861 ms) was lower than that of the target absence condition (2,020 ms), but the difference was nonsignificant, $F(1, 53) = .61$, $p = .437$.

The results from the analyses using the full data from the experimental condition are presented in the Appendix. They were consistent with those based on the first 60 data sets of that condition.

Discussion

The present finding showed that presenting fewer studied probes increased, rather than decreased, the false recognition rate of the CLs. This result clearly supported the dud-effect account over the memory interference or spread-activation account and also contradicted the findings from several previous studies (Coane & McBride, 2006; Diez et al., 2004; Dodd et al., 2006; Marsh et al., 2004) which showed either no effect or an adverse effect of increasing the number of studied probes at test. What could be the sources of the different findings? Dodd et al. (2006), as Jou et al. (2018) and we did in this study (despite the 20-s counting task between study and test), gave the test immediately after subjects studying each list of words, and found no test priming effect, whereas other aforementioned studies administered the test after multiple lists were
studied, and found test priming effects (typically when response was speeded). Could the different findings on test priming effects have resulted from different delays between studying and testing across these studies due to single- versus multiple-lists learning/testing methods? Since more verbatim memory could be retained after a short retention interval than a long interval, it seemed to be a reasonable explanation for the different findings between Dodd’s and other related studies. But it cannot explain the discrepant findings between the present and these related studies. The present study showed that reducing the number of related targets raised the CL false recognition level, whereas those previous studies showed that increasing the number of related targets raised CL false recognition level. Both produced an effect, only in the opposite directions. If the study-test interval rather than the composition of the test list had driven the CL false recognition level, our study should have shown no or very low level of false recognition (as did Dodd et al., 2006). Instead, our experimental condition produced a CL false recognition rate of over .70 which was higher than found in studies that used multiple-lists learning/testing paradigm (e.g., Coane & McBride, 2006; Dewhurst et al., 2011). The present finding was consistent with Jou et al.’s (2018) in that both demonstrated a kind of “reverse test-induced priming effect” (but with the 2018 study manipulating the distractors and the present study manipulating the targets). Another question to be answered is why the dud manipulation affected only the recognition of the CL but not the target, as the hit rates were not significantly different between the control and the experimental conditions. We do not know the answer. Our speculation is that true memory is more resistant to contextual influences than false memory because true memory is based on sensory-perceptual traces, while false memory is constructed from contextual cues. This interpretation appears consistent with the findings of several studies in which warning had an effect in reducing DRM false memory (albeit not remarkably) but had no effect at all (Neuschatz et al., 2001) or a much smaller effect (Gallo et al., 1997; McDermott & Roediger, 1998) on the hit rate of studied words.

Most false memory theories posit that spreading activation is an automatic, involuntary process in generating false memory (McDermott & Watson, 2001; Roediger et al., 2001a, b) and that exposure to an increasing number of related concepts leads to a stronger activation (Jou et al., 2017; Robinson & Roediger, 1997; Roediger et al., 2001a, b). The present result showed that presenting three target probes produced a lower level of false recognition than presenting zero or one target probe. Thus, the automatic spreading activation notion cannot explain this finding. Instead, the result suggested that a slower, judgmental, deliberate process played an important role in the inflated false recognition of the CL. The generally slower responses and especially the longer FA and correct rejection RT of the CLs were consistent with this interpretation. This is contrary to the above-noted findings that allowing more time during retrieval reduces the FA rate of the CLs (or speeding up responses increased FA rate). However, could the slower FA response to the CL in the experimental condition be a result of a lower CL activation level in the experimental than in the control condition? Marsh and Dolan (2007) by testing the CL after 0, 3, or 6 related studied probes found an RT priming effect of number of related test probes on the FA speed to the CL. The higher FA rate associated with a slower response in the present experimental condition was not consistent with that interpretation. If the activation level of the CL in the experiment condition was lower, then, although the RT would be slower, its FA rate should also be lower (Jou et al., 2017), not higher as observed.

But what is the underlying process producing the dud-effect? One account proposed to explain the dud effect is the contrast hypothesis (Charman et al., 2011; Windschitl & Chambers, 2004) which suggests that each comparison between a low-similarity dud with the high-similarity CL increases the confidence in the CL being a target. Another account is the averaged-residual hypothesis (Windschitl & Chambers, 2004) which says that instead of making many pairwise comparisons between the CL and each dud, people assess the averaged supporting evidence provided by the nonfocal alternatives as a whole relative to the evidence of the focal one (equivalent to the CL in the present context). Since the averaged residual evidence is lower for the duds than for the non-duds, the duds increase the subjective likelihood of the focal item being the target.

A related but different account for the dud effect is the range-frequency theory (Parducci, 1965; Wedell, 2008). According to that theory, in a two-category range with two uneven stimulus frequencies, “there is a response tendency toward equal use of the judgment categories so that prediction of the judgments depends in part upon the frequencies with which stimuli are presented from each part of the range” (Parducci, 1965, p. 418). When the frequencies of the two categories of stimuli (duds versus non-duds) are far from equal (the duds far outnumber the CL), the equal response tendency tends to increase the positive response frequency for the CL to reduce the frequency unequalness of the
two stimulus categories. The purpose of the present study is not to adjudicate which of the above accounts is the true cause of the dud effect. The important point we want to make is that memory interference or spreading activation cannot account for the false recognition difference between the control and experimental conditions of the present study. In fact, the RT finding suggests that a deliberate, metacognitive judgmental process similar to a monitoring process (which is normally supposed to lower the false memory) may likely have brought about the higher false memory level in the experimental condition.

Can the results be accounted for by subjects simply adopting a lower recognition decision criterion in the experimental than in the control condition? The result pattern was not consistent with this interpretation, either. First, when the target occurrence rate is lower as in the experimental condition, people are supposed to adopt a stricter decision criterion than when the target rate is higher (Hirshman & Henzler, 1998), and as a result should make fewer, not more FAs. Second, if an overall lower recognition criterion were adopted in the experimental condition, both the hit and the FA rates would be increased. However, in both the partial and full data analyses, the hit rates of the experimental condition were numerically, though not statistically, lower than that in the control condition. Also, the equal distractor rejection rates across the two conditions were also inconsistent with the account of adopting a lower criterion in the experimental condition. Instead, the inflated FA rate seemed to be specifically associated with the CL.

Neither were the RT data consistent with a lower-criterion interpretation of the inflated FA rate to the CL in the experimental condition. A lower decision criterion is typically associated with a faster RT than is a stricter criterion (Jou et al., 2016). The RT for the FA to the CL in the experimental condition was higher than in the control condition, suggesting that subjects might generally be more cautious in judging the experimental-condition CL than the control-condition CL. Previous research showed that when the FA rate of the CL increased, its RT decreased (Jou et al., 2017). The present pattern of an increased RT accompanied by an increased FA rate of the CLs in the experimental condition suggested that the inflated rate of acceptance of the CL was likely based on some slow, deliberate, metacognitive process rather than an automatic, spontaneous process. According to Jacoby (1991), false memory is the product of a fast, automatic activation process, but a slower, controlled monitoring mechanism can be invoked to put this automatic activation process in check. The present study has revealed that this slow, monitoring-like process can also generate more false recognition when the majority of probes in the test is highly implausible.

Appendix

Analyses with the full experimental-condition data

The same response and RT analyses as done for the first 60 subjects’ data were conducted for the full data (of the 132 subjects) of the experimental condition. The overall ANOVA with condition and probe type as factors showed that the mean RT of the experimental condition (1,922 ms) was significantly longer than that of the control condition (1,583 ms), $F(1, 190) = 5.08, p = .025, \eta^2_p = .026$. The main effect of probe type was also significant, $F(2, 380) = 67.43, p < .0001, \eta^2_p = .262$. A Newman–Keuls post hoc analysis indicated that the three means were all significantly different from each other. These RT means are presented in Fig. 2.

![Fig. 2 Mean RTs as a function of condition (experimental vs. control) and probe type using all 132 data sets of the experimental condition](image-url)

The condition by probe type interaction was significant, $F(2, 380) = 3.30, p = .038, \eta^2_p = .017$. The interaction derived from the RT differences between the experimental and control conditions being larger for the CLs and the targets than for the distractors. Thus, the pattern of the results was closely similar to that from the first 60 data sets of the experimental condition.
Additional full-data analysis results are presented in Appendix Table 1.

| Table 1 | Analysis results from full experimental-condition (132 subjects’) data: Means and test statistics |
|---------|---------------------------------|
| **Response Rate** | | |
| Dependent Measure | Control | Experimental | $F$ | $df$ | $p$ | $\eta^2_p$ |
| FA to CL | .562 | .728 | 16.33 | 1, 190 | <.0001 | .079 |
| Hit Rate | .838 | .800 | 1.72 | 1, 190 | =.191 | |
| CR to Distr | .965 | .957 | .34 | 1, 190 | =.558 | |
| Target Pres/Abs on FA to CL | Target Presence | Target Absence | $F$ | $df$ | $p$ | $\eta^2_p$ |
| Target Presence | .721 | .736 | .33 | 1, 131 | =.565 | |
| Target Absence | | | | | | |
| **Response Time** | | | | | | |
| Control | 1,471 ms | 1,978 ms | 10.38 | 1, 185 | = .002 | .053 |
| CR to CL | 2,536 ms | 3,202 ms | 4.73 | 1, 148 | = .031 | .031 |
| Target Pres/Abs on FA to CL | Target Presence | Target Absence | $F$ | $df$ | $p$ | $\eta^2_p$ |
| Target Presence | 1,901 ms | 1,996 ms | .67 | 1, 120 | = .414 | |
| Target Absence | | | | | | |

CR correct rejection, Distr distractor, Pres presence, Abs Absence

Acknowledgments. We thank two anonymous reviewers for their constructive feedbacks. Correspondence can be sent to Jerwen Jou, Department of Psychological Science, University of Texas Rio Grande Valley, 1202 West University Drive, Edinburg, TX 78539-2999, email: jerwen.jou@utrgv.edu

Data availability. Data are available at: https://figshare.com/articles/dataset/DRM_dataset/1695596.

References

Charman, S. D., Wells, G. L., & Joy, S. W. (2011). The Dud effect: Adding highly dissimilar fillers increases confidence in lineup identifications. *Law and Human Behavior, 35*, 479–500.

Coane, J. H., & McBride, D. M. (2006). The role of test structure in creating false memories. *Memory & Cognition, 34*(5), 1026–1036.

Collins, A. M., & Loftus, E. (1975). A spreading activation theory of semantic memory. *Psychological Review, 82*, 407–428.

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology, 55*, 75–84.

Cris, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language, 64*, 316–326.

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

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452–478.

Dewhurst, S. A., Knott, L. M., & Howe, M. L. (2011). Test-induced priming impairs source monitoring accuracy in DRM procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1001–1007.

Diez, E., Fernandez, A., & Alonso, M. A. (2004). Activation at test contributes to false recognition in fast response conditions. Poster at the 45th Annual Meeting of the Psychonomic Society, Minneapolis, MN.

Dodd, M. D., Sheard, E. D., & MacLeod, C. M. (2006). Re-exposure to studied items at test does not influence false recognition. *Memory, 14*, 115–126.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical science. *Behavior Research Methods, 39*(2), 175–199.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*, 833–848.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). The dual-alternative effect in memory for associations: Putting confidence into local context. *Psychonomic Bulletin & Review, 4*, 271–276.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–65.

Hanczakowski, M., Zawadzka, K., & Higham, P. A. (2014). The dual-alternative effect in memory. *Journal of Memory and Language, 70*, 513–541.

Hintzman, D. L. (1988). Judgment of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551.

Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science, 9*(1), 61–65.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Jou, J., Flores, S., Cortes, H. M., & Leka, B. G. (2016). The effects of weak versus strong relational judgments on response bias in two-alternative-forced-choice recognition: Is the test criterion-free? *Acta Psychologica, 167*, 30–44.

Jou, J., Arredondo, M. L., Li, C., Escamilla, E. E., & Zuniga, R. (2017). The effects of increasing semantic-associate list length on the Deese-Roediger-McDermott false recognition memory: Dual false memory process in retrieval from sub- and superspan lists. *Quarterly Journal of Experimental Psychology, 70*(10), 2076–2093.

Jou, J., Escamilla, E. E., Torres, A. U., Ortiz, A., Perez Jr., M., & Zuniga, R. (2018). Why seemingly more difficult test conditions produce more accurate recognition of semantic prototype words: A recognition memory paradox? *Consciousness and Cognition, 63*, 239–253.
Marsh, E. J., & Dolan, P. O. (2007). Test-induced priming of false memories. Psychonomic Bulletin & Review, 14, 479–483.
Marsh, E. J., McDermott, K. B., & Roediger, H. L. (2004). Does test-induced priming play a role in the creation of false memories? Memory, 12, 44–55.
McDermott, K. B., & Roediger, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. Journal of Memory and Language, 39, 508–520.
McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. Journal of Memory and Language, 45, 160–176.
Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. Memory, 9(1), 53–71.
Parducci, A. (1965). Category judgment: A range-frequency model. Psychological Review, 72(6), 407–418.
Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and false recognition. Psychological Science, 8(3), 231–237.
Roediger, H. L. (1996). Memory illusions. Journal of Memory and Language, 35, 76–100.
Roediger, 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, H. L., Balota, D. A., & Watson, J. M. (2001a). Spreading activation and the arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), The nature of remembering: Essays in honor of Robert G. Crowder (pp. 95–115). American Psychological Association.
Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001b). Factors that determine false recall: A multiple regression analysis. Psychonomic Bulletin & Review, 8, 385–407.
Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effective from memory. Psychonomic Bulletin & Review, 4, 145–166.
Underwood, B. J. (1965). False recognition produced by implicit verbal responses. Journal of Experimental Psychology, 70, 122–129.
Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. Journal of Experimental Psychology: Human Learning and Memory, 104, 442–452.
Wedell, D. H. (2008). A similarity-based range-frequency model for two-category rating data. Psychonomic Bulletin & Review, 15(3), 638–643.
Wickens, D. D. (1972). Characteristics of word encoding. In A. W. Melton & E. Martin (Eds.), Coding processes in human memory (pp. 192–215). Wiley.
Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. Journal of Verbal Learning and Verbal Behavior, 2, 440–445.
Windschitl, P. D., & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(1), 198–215.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.