Chapter 25

Associative Memory Illusions

Henry L. Roediger, III, and David A. Gallo

We usually trust our memories. If we get into a discussion with a friend or family members about "what really happened" in some disputed past event, we believe our own memories and not our friend's. But what if we are wrong? Can we remember events differently from the way they really happened? Can we remember events that never happened at all? The answer turns out to be *yes* to both questions.

Distortions of memory arise from many causes. Several types of memory illusions reviewed in this volume are created from external sources. In recollecting some target event from the past, people will often confuse events that happened before or after the target event with the event itself. These confusions are examples of proactive and retroactive interference, respectively (see Chapter 23). On the other hand, the illusion described in this chapter involves the remembering of events that never actually occurred. This erroneous information is internally created by processes that would otherwise lead to good memory for actual events. As such, these errors are part and parcel of the natural memory process, and they are extremely difficult to avoid. Although most of the research reviewed here involves a tightly-controlled laboratory paradigm using word-lists, we believe (and will cite evidence to support) the claim that similar processes occur whenever people try to comprehend the world around them--reading a newspaper or novel, watching television, or even perceiving scenes with little verbal encoding at all (Roediger & McDermott, 2000b).

THE ASSOCIATIVE TRADITION

From Aristotle to computational models, scholars have always assumed that the mind is fundamentally associative in nature (see Roediger, McDermott, & Robinson, 1998). In many theories, associations are viewed as a powerful positive force to support remembering--the stronger the associative bond between two elements, the more probable is retrieval of the second element when given the first as a cue. The idea that associative connections might have a dark side--that they may lead to errors of memory--has hardly ever been considered. However, the point of this chapter is that memory distortions can indeed be induced by associative means.

As far as we know, this idea was first suggested, quite offhandedly, in a paper by Kirkpatrick (1894). He was interested in whether items presented as visual objects were better retained than those presented as words, but his side observations are of interest for present purposes and worth quoting (p. 608):

> About a week previously in experimenting upon mental imagery I had pronounced to the students ten common words … it appears that when such words as “spool,” “thimble” and “knife” were pronounced many students at once thought of “thread,” “needle,” and “fork,” which are so frequently associated with them. The result was that many gave those words as belonging to the list. This is an excellent illustration of how things suggested to a person by an experience may be honestly reported by him as part of the experience.

The process described by Kirkpatrick is the topic of this chapter, how items associated to presented items are often actually remembered as having been overtly presented (rather than inferred covertly). Underwood (1965) did relevant research, but his effect was quite small, and the technique presented in the next section produces much more robust findings. Indeed, false recognition can sometimes be more likely than true recognition when elicited by this newer technique. A simplified version of such an experiment that can be used as a classroom demonstration is described in Text box 25.1 and discussed in the next section.

(Insert Text box 25.1 about here)

SAMPLE EXPERIMENT: THE DRM PARADIGM

Roediger and McDermott (1995) adapted a paradigm first used by Deese (1959) for a somewhat different purpose. The paradigm – the one described in Text box 25.1--produces a very strong associative memory illusion and (owing to a suggestion by Endel Tulving) is now called the DRM paradigm (for Deese-Roediger-McDermott). The paradigm and variants of it are frequently used as a straightforward technique for gaining measures of both veridical and false memories using both recall and recognition techniques. We describe here a somewhat simplified version of Experiment 2 in Roediger and McDermott (1995).

Typical method

A set of 24 associative lists were developed, each list being the 15 strongest associates to a nonstudied word, as found in word association norms. These norms are based on a free association task, in which participants were presented a stimulus word (e.g., *rough*) and told to generate the first word that comes to mind. To create each of our lists, we took the 15 words that had been elicited most often by the stimulus word (e.g., the words *smooth, bumpy, road, tough, sandpaper, jagged, ready, coarse, uneven, riders, rugged, sand, boards, ground,* and *gravel*; see Roediger, Watson, McDermott, & Gallo, 2001, for a set of 55 lists with normative false recall and recognition data). These study lists were presented to new participants, and their memory was subsequently tested. Critically, the stimulus word (*rough* in this case) was never studied by these participants. Our interest centered on the possible false recall or false recognition of this critical word. If a participant were like a computer or tape recorder, recording and retrieving the words perfectly, one would not expect such systematic memory errors.

Thirty undergraduate participants heard 16 of the 24 word lists, one word at a time. Participants recalled 8 of the lists immediately after their presentation (with two minutes provided for recall), and they were instructed not to guess but only to recall the words they were reasonably sure had been in the list. They performed arithmetic problems for two minutes after each of the other 8 lists. Shortly after all 16 of the lists had been presented in this way, participants took a yes/no recognition test that covered all 24 lists. That is, they saw words one at a time and said yes or no as to whether or not the item had been studied in the list. Because only 16 of the 24 lists had been studied, items from the other 8 lists served as lures or distractors to which participants should respond "no" (it was not on the list). The recognition test was composed of 96 words, with 48 having been studied and 48 new words. Importantly, 16 of these new words were the critical missing lures (words like *rough*) that were strongly associated to the studied words.

After each word that they recognized as having been in the list (the ones judged “yes”), participants made a second judgment. They were asked to judge whether they remembered the moment of occurrence of the word in the list, say by being able to remember the word before or after it, what they were thinking when they heard the word, or some other specific detail. This procedure is called a *remember* judgment and is thought to reflect the human ability to mentally travel back in time and re-experience events cognitively. If they were sure the word had been in the list, perhaps because it was highly familiar, but could not remember its specific moment of occurrence, they were told to make a *know* judgment. The remember/know procedure was developed by Endel Tulving in 1985, and has since been used by many researchers in order to measure the phenomenal basis of recognition judgments (see Umanath & Coane, 2020, for a critical analysis of the distinciton).

Typical results

Let us consider the immediate free recall results first. Recall of list items followed the typical serial position curve, with marked primacy and recency effects reflecting good recall at the beginning (primacy) and end (recency) of the list. Consider next false recall of the critical nonpresented word such as *rough* (in our sample list used above). Despite the fact that recall occurred immediately after each list and participants were told not to guess, they still recalled the critical nonpresented item 55% of the time! In this experiment, recall of the critical nonpresented item was actually higher than recall of the items that were presented in the middle of the list. In other studies, the probability of recall of critical items often approximates recall of items in the middle of the list, with the particular outcome depending on such factors as presentation rate of the lists (1.5 sec in this study) and whether the lists are presented auditorily or visually. The important point is that false recall was very high.

The recognition test also revealed a powerful associative memory illusion. The basic data are presented in Figure 25.1. Shown in the two panels are data from the 8 lists that were studied and recalled (the right panel) and from the 8 lists that were studied but not previously recalled (the left panel). Within each panel, the left bar shows veridical or true recognition (the hit rate) of items actually studied, whereas the right bar shows false recognition of the critical lures like *rough* (the critical-lure false alarm rate). The false alarm rates to the items from the 8 nonstudied lists that appeared on the test are given in the figure caption. Finally, each bar is divided into a dark portion (items called old and judged to be *remembered*) and a white portion (items called old and judged to be *known)*.

(Insert Figure 25.1 about here)

Figure 25.1 shows the very large false recognition effect that is typical of the DRM paradigm. For example, for lists that were studied and had been recalled, participants recognized 79% of the list words as old and said they remembered 57% of the words. Nearly three-quarters of the words called old were judged to be remembered (i.e., 57/79 = 72%). Surprisingly, the data for the critical lures (words that were not actually presented) were practically the same! Participants recognized 81% as old and even judged 58% *remembered*. Thus, the participants recalled words that were never presented, and, just like for studied words, 72% of the words judged old were remembered (58/81 = 72%). So, in the DRM paradigm, the level of false recognition and false remembering is about the same as veridical recognition and remembering of list words. People recall and remember events that never happened. The situation is much the same for the lists that were studied but not recalled, with false recognition of critical lures being as high as (or even higher than) veridical recognition of list words. Note, however, that remember judgments were lower on the recognition test (for both kinds of items) when the lists had not been recalled. In some ways, the data in the left panel show effects of recognition that are "purer" in that they were not affected by prior false recall. Nonetheless, striking levels of false recognition and "remember" judgments were obtained. And note that the act of recall boosted both accurate and false recall, an example of the effect of retrieval practice on later retention (see McDermott, 2021). Practicing retrieval boosts recall of both veridical and false memories (see also McDermott, 2006).

Discussion

Perhaps because the effect is so robust, a skeptical reaction after learning about the effect is disbelief: Participants are obviously not trying to remember at all. Instead, this reasoning continues, participants are faced with too many words to remember, and so make educated guesses as to which words were presented. In particular, they realize that the lists consist of associated items, so they infer that critical items (which are associated to the study lists) were also presented. Miller and Wolford (1999) formalized this sort of decision process in terms of a liberal criterion shift to any test word that is perceived as related to the study list (i.e., the critical items). This model was primarily directed at false recognition, although a generate/recognize component was included to account for false recall. In either case, it is assumed that participants try to capitalize on the related nature of the lists (via a liberal guessing strategy for related items), in the hopes of facilitating the correct recognition of studied words.

Gallo, Roediger, and McDermott (2001) directly tested this account by informing participants about the illusion and telling them to avoid false recognition of nonstudied but related words. The critical condition was when participants were warned after studying the lists but just before taking the recognition test, thereby precluding a liberal guessing strategy for related items. The results were straightforward: Warning participants between study and test had negligible effects on false recognition (relative to a no-warning control condition), even though other conditions revealed that warned participants were trying to avoid false recognition. This pattern also was obtained in false recall (e.g., Neuschatz, Payne, Lampinen, & Toglia, 2001). Gallo et al. (2001) reasoned that warned participants did not adopt a liberal strategy to related items because, after all, they were trying to avoid false alarms to these items. Thus, robust false memory effects following warnings were not due to such strategic decision processes alone, but instead were due to processes that are inherent in the memory system. This conclusion is bolstered by the finding that participants will often claim that these false memories are subjectively detailed and compelling (as reviewed below).

In sum, the DRM paradigm is one of the most potent memory illusions ever studied. As noted above, similar associatively-based errors have been obtained using a wide variety of materials, including pictures, sentences, and stories, although these errors are usually not as frequent as those observed in the DRM paradigm (see Roediger & McDermott, 2000a, 2000b, for an overview). In general, any set of materials that strongly implies the presence of some object or event that is not actually presented lends itself to producing false recall and false recognition of the missing but implied event (cf. Chapter 26 on the misinformation effect). Why, then, are the false memories produced by the DRM paradigm so robust?

There are several answers to this question, but we will concentrate on the most critical one: the number of associated events that are studied. The DRM paradigm, in contrast to most other memory illusions, relies on the presentation of multiple associates to the critical nonstudied word, thereby taking full advantage of the power of associations. Robinson and Roediger (1997) directly examined the effect of number of associated words on false recall and false recognition. In one experiment, they presented participants with lists of 15 words prior to recall and recognition tests, but the number of words associated to a critical missing word was varied to be 3, 6, 9, 12, or 15. Increasing numbers of associated words steadily increased false recall of the critical nonpresented word, from 3% (with 3 words) to 30% (with 15 words). Thus, even though the total number of words studied was the same in all conditions, the number of studied associates to the critical word had a considerable influence on the strength of this memory illusion. We discuss the theoretical implications of this finding in the next section.

THEORIES AND DATA

In this section we consider the processes that may be involved and how they might interact to give rise to the associative memory illusion. This discussion is divided into two main sections: processes that cause the effect and opposing processes that reduce the effect. Our goal is not to exhaustively review all of the DRM findings – that would be well beyond the scope of this chapter. Indeed, Gallo (2006) wrote a book-length review of the first 10 years of DRM research, and 15 years of additional research have ensued since then (see Huff, Bodnar, & Fawcett, 2015, and Pardilla-Delgado & Payne, 2017, for more recent reviews). Rather, our goal in the present chapter is to highlight the main theoretical issues and discuss those DRM findings that we feel critically inform these issues. In many instances, more than one group of researchers reported relevant findings, but for brevity we cite only one or two findings to illustrate the point.

Processes that cause the effect

The dominant theories of the DRM effect fall into two classes: *association-based* and *similarity-based*. These classes differ in the types of information or representation that is proposed to cause false remembering (and in terms of the processes that allegedly give rise to these representations). Nevertheless, these theories are not mutually exclusive, and evidence suggests that both types of mechanism make a unique contribution to the effect. We discuss each in turn, followed by a brief consideration of attribution processes that may contribute to the subjectively detailed nature of associative false memories.

Association-based theories

According to the association-based theories, a preexisting representation of the critical nonpresented word becomes activated when its associates are presented. Thus, presenting *bed*, *rest*, *awake* etc. activates the mental representation of the word *sleep*. Under this theory, false remembering occurs when the participant mistakes associative activation with actual presentation, which can be conceptualized as a reality monitoring error (Johnson, Hashtroudi, & Lindsay, 1993). That is, did the item really occur during the list presentation, or did I just think about it? This theory is similar to Underwood's (1965) classic idea of the implicit associative response (IAR), and is consistent with Deese's (1959) finding that the degree of association between the list words and the critical nonpresented word (dubbed Backward Associative Strength, or BAS) was highly predictive of false recall. That is, the more items in the list have the critical item as an associate (which is what BAS measures), the more likely the list is to produce false recollection.

Deese's (1959) finding was replicated and extended by Roediger, Watson, McDermott, and Gallo (2001), who reported that BAS predicted most of the variance in false recall (among several candidate variables) using multiple regression analysis. Roediger et al. (2001) interpreted this relationship as evidence for associative activation. The notion is that associates activate the lexical representation of the critical word, and this activation supports the thought of the item on a recall test. They also found that BAS was related to false recognition, suggesting that activation might be a common cause of false recall and false recognition, although the differences in recognition tend to be somewhat smaller than those found in recall (see Gallo & Roediger, 2002). The aforementioned list-length effect (e.g., Robinson & Roediger, 1997) is also consistent with an associative activation mechanism: Increasing the number of associates studied increases associative activation, and hence increases false recall and recognition.

Two obvious questions concern the form of this activation (conscious or nonconscious?) and when it occurs (study or test?). The fact that false recall and false recognition occur even with very rapid study presentation rates (under 40 ms per item, or less than a second per list) suggests that conscious thoughts of the critical item during study are not necessary to elicit false memory (see McDermott & Watson, 2001, for recall evidence, and Cotel, Gallo & Seamon, 2008, for recognition evidence). This is consistent with semantic priming models, which suggest that associative activation at study can automatically spread from one word node to another (see Roediger, Balota, & Watson, 2001). However, just because conscious thoughts of the critical item may not be necessary to elicit false remembering does not imply that they do not occur at the relatively slower presentation rates (e.g., 1-2 s per item) that are typically used in the paradigm. At more standard rates, overt rehearsal protocols indicate that participants often think of the critical item during study, and the frequency of these thoughts predicts subsequent false recall (e.g., Goodwin, Meissner, & Ericsson, 2001).

Additional evidence that associative activation occurs at study has been obtained using implicit tests. After presenting participants with several DRM lists, McDermott (1997) found priming for the critical items on several implicit memory tests, and these effects have since been replicated and extended to other priming tasks (e.g., Meade, Hutchison, & Rand, 2010). McDermott (1997) argued that such priming was due to lexical activation of the critical item at study.

Similarity-based theories

Within the second class of theories, the core proposal is that DRM false remembering is caused by similarity between the critical item and the studied items, as opposed to associative activation of the critical item. These theories have primarily been used to explain false recognition. For instance, the fuzzy trace theory of memory representations (e.g., Brainerd, Wright, Reyna, & Mojardin, 2001) postulates that studying a list of associates results in the formation of two types of memory traces. Verbatim traces represent detailed, item-specific information, whereas gist traces represent the more general thematic characteristics of the lists based on meaning of the words. At test, words that are consistent with the gist of the list (such as the critical item) will be highly familiar, and hence falsely remembered. A different similarity-based account was developed by Arndt and Hirshman (1998), as an extension of exemplar-based models of memory. Under their proposal, a separate "gist" representation need not be encoded. Instead, each studied item is encoded as a set of sensory and semantic features. At retrieval, the similarity between the features of the critical item and the encoded features will make this item familiar, and lead to false remembering.

Despite these differences, both of these similarity-based theories explain DRM false recognition via familiarity caused by semantic similarity, and neither theory appeals to activation of the critical item through associative links (see Brainerd et al., 2008). This last point poses important constraints on these theories. Without positing some sort of item-specific activation of the critical item, it is difficult to understand how these theories would explain the generation of this item on a recall test or on perceptually driven implicit memory tests (such as word stem completion). Theories based entirely on semantic similarity also have difficulty explaining why categorized lists (e.g., different pieces of furniture) are less effective than DRM lists at eliciting false memory effects (Pierce, Gallo, Weiss, & Schacter, 2005). Categorized lists are strongly thematic, but tend to have weaker associations than DRM lists.

Perhaps the strongest evidence that similarity-based processes might be involved in addition to associative activation are the effects of retention interval. It has been found that true recall decreases more over a delay than false recall (e.g., Toglia et al., 1999). Illustrative data from Toglia et al. (1999) are presented in Figure 25.2. True recall declined rapidly over a three-week retention interval, whereas false recall persisted at high levels. Fuzzy trace theory can account for such results because it holds that gist traces are more resistant to forgetting than verbatim traces. As a result, memory for list items (which is supported more by verbatim traces) decreases at a more rapid rate than memory for critical items (which is supported more by gist traces) as retention interval is increased.

(Insert Figure 25.2 about here)

Associative-based theories cannot account for such effects without additional assumptions. In the strongest form of these theories, the critical item would be activated multiple times at study and rehearsed like a list item. To the extent that the critical item is encoded like a studied item, the two should have similar forgetting functions (especially when initial levels of true and false remembering were matched). In return, it is unclear how a similarity-based mechanism could account for the powerful relationship between associative strength and false remembering. For example, many of the words in the *whiskey* list seem to converge on the meaning of that word (e.g., *drink, drunk, beer, liquor*, etc.) just as words in the *window* list converge on its meaning (e.g., *door, glass, pane, shade*, etc.). Nevertheless, these lists greatly differ in associative strength (0.022 vs. 0.184), and in turn, they elicit dramatically different levels of false recall (3% vs. 65%; for additional discussion see Gallo & Roediger, 2002; Roediger et al., 2001). In sum, it appears that both associative activation and semantic similarity play a role.

Another line of evidence is relevant to the discussion. Activation theories postulate two different types of activsation: lexical (based on the form and sound of the word) and semantic (based on word meaning). As already noted, McDermott's (1997) implicit memory results were credited to lexical activation rather than semantic activation. In addition, Sommers and Lewis (1999) showed that lists of words phonologically associated with a missing word leads to false recall and false recognition. For example, for a word like *speak*, the list words might include *sneak, sleek, peak, cheek,* etc. Sommers and Lewis found that such lists did create recall of the target word that was not presented in the list (*speak*, in this case). Because phonological associates are not related in meaning to the target word, the assumption is that they activated a lexical level of representation and that activation spreads through the lexical network. This idea is in accord with a theory called the Neighborhood Activation Model (Luce, 1998). The model postulates that phonological associates of a word are represented in space, with some closer and some further away in the "neighborhood" of the associates.

Watson, Balota, and Roediger (2003) asked what would happen if hybrid lists were created, such that half the items converged on meaning of the missing word and half on its phonology or sound. So, for the word *sleep* (omitted from the list), the list words might include *bed, rest, awake*, and *tired* for the semantic associates and *sheep, peep, beep*, and *sleet* for the phonologically similar (rhyming) words. These researchers asked if the false memories that occurred from using such hybrid lists would be additive; that is, would the effects of a combined list with phonological and semantic associates be equal to the sum of presenting the phonological and semantic items when they are presented alone? Watson et al. (2003) expected to find additivity, but found that the hybrid lists created superadditive effects, meaning that the hybrid lists created false memories greater than the sum of the two independent lists that comprised the hybrid lists. That is, they found superadditivity! The activation from both the semantic and phonological levels zeroes in on the missing item and leads to greater false recall. In one study, semantic lists created .34 false recall of the missing item, whereas phonological lists created .14 false recall. By the additivity rule, false recall from hybrid lists should be around .48 (i.e., .34 +.14). However, actual false recall in the hybrid lists was .65, one of the largest false-memory effects ever obtained.

Finley, Sungkhassette, Balota, and Roediger (2017) confirmed the overadditity effect and plotted out the effects of various levels of hybridization (e.g., 13 phonological associates and 3 semantic associates). In general, adding only a few semantic associates to a list of mostly phonological associates produced a large increase in false recall; the same was true of adding a few phonological associates to a semantic list.

The phonological false memory effect and the powerful effect of using hybrid lists to create superadditive false memories are explainable in a straightforward manner by associative activation theories. For hybrid lists, activation of the critical missing item cumulates from spreading activation through both a lexical and a semantic network, converging on the nonpresented associate. On the other hand, fuzzy trace theory, with its exclusive focus on meaningful gist as causing false memory effects, seems unable to account for these effects. However, recently Chang and Brainerd (2021) have extended fuzzy trace theory to account for phonological effects, too.

Fluency-based attributions

Although they can explain false recall and false recognition, neither the associative-based account nor the similarity-based account can explain the perceptually detailed nature of DRM false memories very well. Roediger and McDermott (1995) found that false recognition of critical items was accompanied with high levels of confidence and frequent *remember* judgments. Both of these findings can be explained by thoughts of the critical item at study, but even this account cannot explain more detailed recollections. For instance, when lists are presented by multiple sources (auditory vs. visual, or different voices), participants are often willing to assign a source to critical items that are falsely recognized (Roediger, McDermott, Pisoni, & Gallo, 2004) or recalled (Hicks & Marsh, 1999). Similarly, using the Memory Characteristics Questionnaire (MCQ), participants often claim to recollect specific details about a critical item's presentation at study, such as perceptual features, list position, and personal reactions to the word (e.g., Mather, Henkel, & Johnson, 1997).

The cause of such illusory subjective phenomena is still debated in the literature (see Arndt, 2012), but one possible explanation is a *fluency-based attribution* process. Gallo and Roediger (2003) proposed that, at test, participants imagine having been presented with the critical item at study, perhaps in an effort to determine whether it was presented. This imagination is then mistaken for actual presentation because it is processed more fluently, or more easily, than would have otherwise been expected (cf. Chapter 11 on availability, Chapter 14 on the validity effect, Chapter 15 on the mere exposure effect, and Chapter 26 on the misinformation effect). If the attribution process occurs automatically, or nonconsciously, then the phenomenological experience would be one of remembering (cf. Jacoby, Kelley, & Dywan, 1989).

In addition to imagination, events that were actually studied may provide another source of details for these kinds of false memories. Lampinen, Neuschatz, and Payne (1999) used the term "content borrowing" to describe how features of studied items could be retrieved in a fragmented way, such that these free-floating features in memory could then be mistakenly bound into a false memory for the nonpresented associate. This process could cause a detailed yet false recollection of the nonstudied event, especially if the item is processed fluently. Indeed, more recent work on false memories for nonstudied pictures has demonstrated that the conceptual fluency of an item as well as the availability of fragmented perceptual features in memory can independently drive false recollection effects (Doss, Bluestone, & Gallo, 2016).

Both the associative-based and similarity-based theories predict that processing of the critical word will be enhanced by presentation of the related list, so that a fluency-based attribution process is consistent with either theory. That said, there are some clues that associative activation is uniquely involved (for examples, see Franks, Butler, & Bishop, 2016; Gallo & Roediger, 2002, 2003).

Processes that reduce the effect

So far we have discussed processes that drive the DRM effect. No theoretical account would be complete, though, without considering editing processes that oppose these forces and reduce false remembering. Such processes have been conceptualized as *reality monitoring* under association-based theories (e.g., activation/monitoring theory), and item-specific or verbatim-based editing in similarity-based theories (e.g., fuzzy trace theory).

Some of the earliest evidence that such additional processes are involved comes from presentation manipulations that should not affect associative activation or semantic similarity, but nevertheless influence false remembering. These include presentation format (e.g., switching presentation from words to pictures, which has been found to reduce false recognition; Schacter, Israel, & Racine, 1999) and presentation modality (e.g., switching presentation from auditory to visual, which reduces the DRM effect; Smith & Hunt, 1998). Although more recent work indicates these kinds of manipulations might impact activation in addition to monitoring processes in the DRM task (e.g., Smith & Hunt, 2019), these same manipulations also have been found to reduce false memory effects in source recollection tasks that do not involve the activation of semantic associations (for review, see Gallo, 2013). Thus, there is converging evidence across different tasks that both presentation format and presentation modality can impact monitoring processes.

Other evidence for monitoring processes in the DRM task comes from presentation manipulations that should increase similarity or associative processes, but actually decrease false remembering. These include increasing the number of presentations of the study lists before a recognition test (e.g., Benjamin, 2001), and slowing presentation rate (which has been found to reduce false recall, but not necessarily false recognition; Gallo & Roediger, 2002). To illustrate, consider a presentation-rate study by McDermott and Watson (2001). In those conditions that are relevant here, participants studied DRM lists at a range of visual presentation durations (20, 250, 1000, 3000, and 5000 milliseconds, between-subjects), and took an immediate free recall test after each list. As expected, true recall increased with more study time (0.17, 0.31, 0.42, 0.50, and 0.51). The pattern for false recall was more striking, with an initial increase and an eventual decrease (0.14, 0.31, 0.22, 0.14, and 0.14). The initial increase suggests that, within this range of extremely rapid presentation rates, slowing the duration afforded more meaningful processing and thus enhanced those activation-based or similarity-based processes that drive false recall. In contrast, the eventual decrease suggests that slowing presentation rates also increases item-specific processing of the list items. Apparently, the accrual of this item-specific information eventually reached a point where it began to facilitate monitoring processes that opposed false recall.

Subsequent research indicates that there are different kinds of monitoring or editing processes that influence DRM false remembering, as well as other kinds of false memories more generally (Gallo & Lampinen, 2016). One monitoring process – dubbed the distinctiveness heuristic by Schacter et al. (1999) – relies on the idea of retrieval expectations. According to this idea, making the studied items more memorable or distinctive allows participants to expect richer or more detailed memories at retrieval, effectively setting a more conservative decision criterion that helps them to reject false memories that fail to meet these expectations. Another kind of monitoring process – often called recall-to-reject – occurs when participants realize, during the presentation of the study list, that the critical item is missing (Carneiro & Fernandez, 2013). If they remember this realization at test, then they can avoid falsely remembering the critical item regardless of the distintiveness of the studied information. This type of monitoring is most likely to occur if participants are warned to avoid the DRM illusion prior to the study phase, although of course the standard procedure does not give such warnings.

Neural mechanisms of the effect

We have discussed how both activation/similarity and editing processes may play a role in the DRM illusion. Further support for the distinction between these two opposing processes comes from neuropsychological data. Amnesic patients with varied etiologies (e.g., Korsakoff's or anoxia) tend to show decreased DRM false recognition relative to age-matched controls (e.g., Schacter, Verfaellie, Anes, & Racine, 1998). This decrease implies that damage to medial temporal regions such as the hippocampus (which were the primary, but not the sole areas that were damaged) reduces the likelihood of remembering the associative relations or gist that can cause false remembering. Related effects have been found in particiapants in the early stages of Alzheimer's disease, which also affects medial temporal regions (e.g., Gallo et al., 2006).

In contrast to those effects, patients with frontal lobe lesions showed enhanced DRM false recognition relative to age-matched controls (e.g., Budson, Sullivan, Mayer, Daffner, Black, & Schacter, 2002). The frontal lobes have traditionally been implicated in monitoring processes, suggesting that the elevated levels of false recognition in this population were due to a breakdown in false-memory editing. Advanced aging also can increase susceptibility to the DRM illusion, especially in those older adults with poor frontal-lobe functioning (Butler et al., 2004). This effect has been attributed to a breakdown in the ability to monitor memory, coupled with a spared ability to process semantic associations. Considered as a whole, the data from these different populations nicely illustrate the opposing influences of activation/similarity and editing processes.

Developmental patterns also provide a unique window into these opposing influences. Young children have underdeveloped frontal lobes, and in general, they tend to have difficulty monitoring memory and avoiding false-recognition errors in many memory tasks (e.g., Moore, Lampinen, Bridges, & Gallo, 2020). However, in the DRM task, young children tend to be less susceptible to false memories compared to older children and adults. This effect has been attributed to underdeveloped semantic processing in young children, such as difficulties in connecting and comprehending the semantic associations in DRM lists (e.g., Brainerd, Reyna, & Holliday, 2018). Analogous to patients with damage to medial temporal lobes, children are less likely to “get the gist” than adults, leading to reduced relatedness effects on false memory.

Data from neuroimaging techniques, such as fMRI or EEG, have provided further insights into the neural mechanisms of the DRM effect. Unlike lesion studies, these techniques can separate neural involvement at encoding from retrieval, although findings tend to be highly specific to nuances of the experimental design and analytical techniques. Despite these challenges, a meta-analysis of 34 fMRI studies using the DRM task and related tasks by Kurkela and Dennis (2016) identified two persistent findings that deserve highlighting. The first finding was that brain regions typically associated with the meaningful processing of language (e.g., left prefrontal and left middle temporal gyrus) tend to be activated during the encoding phase, potentially owing to the processing of semantic associations (see McDermott, Watson, & Ojemann, 2005). The second finding was elevated activity in numerous prefrontal regions during false memory retrieval, which were typically attributed to monitoring demands. When considered with the aforemented studies of neuropsychological populations, these neuroimaging studies provide converging evidence for the distintion between activation/similarity and editing processes.

CONCLUSION

Associative memory illusions arise when information from the external world activates internal representations that may later be confused with the actual external events that sparked the association. As we have emphasized, we believe that this process is a general one with wide implications, because such associative activation is a pervasive fact of cognition. To use Jerome Bruner's famous phrase, people frequently go "beyond the information given" in drawing inferences, making suppositions, and creating possible future scenarios. Although these mental activities make us clever, they can also lead to errors when we confuse what we thought with what actually happened. The

DRM paradigm provides a tractable laboratory task that helps open these processes to careful experimental study, and it also provides a rich arena for testing theories of internally-generated false memories.

SUMMARY

- People can falsely remember nonpresented events that are associated to events that occurred.
- Research has identified two sets of factors that are critical for the creation of these types of false memories: activation processes and monitoring processes.
- Activation processes, such as the mental generation of associative information, cause people to believe that the nonpresented event had actually occurred.
- Monitoring processes refer to the strategic editing of these retrieval products, in an effort to reduce false remembering.
- The frequent occurrence of these systematic errors provides important insights into the cognitive mechanisms of memory.

FURTHER READING

For other relevant DRM reviews, see Gallo (2010) and Huff et al. (2015). For some other perspectives see Schacter, Norman, and Koutstaal (1998) and Mitchell and Johnson (2009). Finally, while intriguing new findings are coming out every day (Wang, Otgaar, Santtila, Shen, & Zhou, 2021), it also is good to keep a historical perspective. On that note, see Bruce and Winograd (1998).

REFERENCES

Arndt, J. (2012). False recollection: Empirical findings and their theoretical implications. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 56; pp. 81-124). San Diego: Academic Press.

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language*, *39*, 371-391.

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 941-947.

Brainerd, C. J., Reyna, V. F., & Holliday, R. E. (2018). Developmental reversals in false memory: Development is complementary, not compensatory. *Developmental Psychology, 54*, 1773-1784.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 307-327.

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychonomic Bulletin & Review, 15,* 1035-1053.

Bruce, D., & Winograd, E. (1998). Remembering Deese's 1959 articles: The Zeitgeist, the sociology of science, and false memories. *Psychonomic Bulletin & Review*, *5*, 615-624.

Budson, A. E., Sullivan, A. L., Mayer, E., Daffner, K. R., Black, P. M., & Schacter, D. L. (2002). Suppression of false recognition in Alzheimer's disease and in patients with frontal lobe lesions. *Brain*, *125*, 2750-2765.

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L., III. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review, 11*, 921-925.

Calkins, M. W. (1894). Association studies from the Harvard Psychological Laboratories. *Psychological Review*, *1*, 476-483.

Chang, M., & Brainerd, C. J. (2021). Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language, 119*, Article No. 104210.

Cotel, S. C., Gallo, D. A., & Seamon, J. G. (2008). Evidence that nonconscious processes are sufficient to produce false memories. *Consciousness & Cognition*, *17*, 210-218.

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

Doss, M. K., Bluestone, M. R., & Gallo, D. A. (2016). Two mechanisms of constructive recollection: Perceptual recombination and conceptual fluency. *Journal of Experimental Psychology: Learning, Memory and Cognition, 42*, 1747-1758.

Ebbinghaus, H. (1885/1913). *On memory: A contribution to experimental psychology*. New York, Dover Publishing Co.

Finley, J. S., Sungkhasette, V. W., Balota, D. A., & Roediger, H. L. (2017). Relative contributions of semantic and phonological associates to over-additive false recall in hybrid DRM lists. *Journal of Memory and Language, 93*, 154-168.

Franks, B. A., Butler, K. M., & Bishop, J. (2016). The effects of study order and backward associative strength on illusory recollection: A source-strength effect does not always occur. *Memory*, *24*, 154-164.

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York: Psychology Press.

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

Gallo, D. A. (2013). Retrieval expectations affect false recollection: Insights from a criterial recollection task. *Current Directions in Psychological Science, 22,* 316-323.

Gallo, D. A., & Lampinen, J. M. (2016). Three pillars of false memory prevention: Orientation, evaluation, and corroboration. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory*. Pp. 387-403. New York: Oxford University Press.

Gallo, D. A., & Roediger, H. L., III. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language*, *47*, 469-497.

Gallo, D. A., & Roediger, H. L., III. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition*, *31*, 1036-1044.

Gallo, D. A., Roediger, H. L., III, & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, *8*, 579-586.

Gallo, D. A., Shahid, K. R., Olson, M. A., Solomon, T. M., Schacter, D. L., & Budson, A. E. (2006). Overdependence on degraded gist memory in Alzheimer's disease. *Neuropsychology, 20*, 625-632.

Goodwin, K. A., Meissner, C. A., & Ericsson, K. A. (2001). Toward a model of false recall: Experimental manipulations of encoding context and the collection of verbal reports. *Memory & Cognition*, *29*, 806-819.

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1195-1209.

Hicks, J. L., & Starns, J. J. (2006). The roles of associative strength and source memorability in the contextualization of false memory. *Journal of Memory and Language, 54*, 39–53.

Huff, M. J., Bodnar, G. E., & Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: A meta-analytic review of costs and benefits and their orign in the DRM paradigm. *Psychonomic Bulletin & Review, 22,* 349-365.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.

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

Kurkela, K. A., & Dennis, N. A. (2016). Event-related fMRI studies of false memory: An activation likelihood estimation meta-analysis. *Neuropsychologia, 81*, 149-167.

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review, 6,* 130–135.

Luce, P. A. (1987). The neighborhood activation model of auditory word recognition *The Journal of the Acoustical Society of America, 81*, S1.

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

McDermott, K. B. (2006). Paradoxical effects of testing: Repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Memory & Cognition, 34*(2), 261-267

McDermott, K. B. (2021). Practicing retrieval facilitates learning. *Annual Review of Psychology, 72*, 609-633.

McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review*, *4*, 582-586.

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.

McDermott, K. B., Watson, J. M., & Ojemann, J. G. (2005). Presurgical language mapping. *Current Directions in Psychological Science, 14,* 291-295.

Meade, M. L., Hitchison, K. A., & Rand, K. M. (2010). Effects of delay and number of related list items on implicit activation for DRM critical items in a speeded naming task. *Journal of Memory and Langauge, 62*, 302-310.

Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, *106*, 398-405.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*, 638–677.

Moore, K. N., Lampinen, J. M., Bridges, A. J., & Gallo, D. A. (2020). Developmental trends in children's use of different monitoring processes to avoid false memories. *Cognitive Development, 55*, 100911.

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*, 53-71.

Pardilla-Delgado, E., & Payne, J. (2017). The Deese-Roediger-McDermott (DRM) task: A simple cognitive paradigm to investigate false memories in the laboratory. *Journal of Visualized Experiments*, *119*, 54793.

Pierce, B. H., Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2005). The modality effect in false recognition: Evidence for test-based monitoring. *Memory & Cognition*, *33*, 1407-1413.

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

Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and the arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95-115). Washington, D. C.: American Psychological Association Press.

Roediger, H. L., III, & 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., III, & McDermott, K. B. (2000a). Distortions of memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 149-162). New

York: Oxford University Press.

Roediger, H. L., III, & McDermott, K. B. (2000b). Tricks of memory. *Current Directions in Psychological Science*, *9*, 123-127.

Roediger, H. L., III, McDermott, K. B., Pisoni, D. B., & Gallo, D. A. (2004). Illusory recollection of voices. *Memory, 12*, 586-602.

Roediger, H. L., III, McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187-245). Hove, England: Psychological Press.

Roediger, H. L., III, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385-407.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Rememering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*, 773-786.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1-24.

Schacter, D. L., Verfaellie, M., Anes, M. D., & Racine, C. (1998). When true recognition suppresses false recognition: Evidence from amnesic patients. *Journal of Cognitive Neuroscience*, *10*, 668-679.

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review, 5*, 710-715.

Smith, R. E., & Hunt, R. R. (2020). When do pictures reduce false memory? *Memory &*

Cognition, 48, 623-644.

Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language, 40*, 83-108.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, *7*, 233-256.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1-12.

Umanath, S., & Coane, J. H. (2020). Face validity of remembering and knowing: Empirical consensus and disagreement between participants and researchers. *Perspectives on Psychological Science*, *15*(6), 1400-1422.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, *70*, 122-129.

Wang, J., Otgaar, H., Santtila, P., Shen, X., & Zhou, C. (2021). How culture shapes collective false memory. *Journal of Applied Research in Memory and Cognition*, *10*, 24-32.

Watson, J. M., Balota, D. A., & Roediger, H. L. (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory and Language, 49*, 95–118.

TEXT BOX

Text box 25.1

Classroom demonstration

This demonstration can be used to create false memories in only a few minutes. For best results, participants should not be told that the demonstration is on false memories until after the experiment. We will suggest two variations on this theme after this demonstration so that you may compare two other interesting conditions.

Material

The material consists of four lists with 15 words each that are all associated to a critical, but not included target word.

> List 1: *bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap, peace, yawn, drowsy.*
>
> List 2: *door, glass, pane, shade, ledge, sill, house, open, curtain, frame, view, breeze, sash, screen, shutter.*
>
> List 3: *nurse, sick, lawyer, medicine, health, hospital, dentist, physician, ill, patient, office, stethoscope, surgeon, clinic, cure.*
>
> List 4: *sour, candy, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, heart, cake, tart, pie.*

The critical target words are *sleep*, *window*, *doctor*, and *sweet*, respectively.

Procedure

The experimenter tells the participants that this will be a memory demonstration, and that they should have a piece of scrap paper and a pen ready for the memory test. The experimenter then tells them that s/he will read lists of words, and that they should try to remember these words. The participants should not be allowed to write the words down as they are being read. The experimenter then reads the first list, at a steady rate of one word every 1 to 2 seconds. After the final word, the participants are asked to write down as many words as they can remember, in any order, without guessing. Participants usually take less than a minute to recall each list. This procedure is then repeated for the next three lists.

Analysis

After the final list is recalled, the experimenter counts separately for each list the number of participants (by having them raise their hands or by tallying the recall sheets) who recalled the critical word. As these critical associates were never presented, their recall represents false memories.

Variations

The demonstration above is quite simple. Here are two variations on the same idea. Using the same condition described above as the control condition, test another group of participants but explain the phenomenon to them *before* they are presented with the lists. That is, tell them that they are to be given a list of words that is intended to make them think of another word and to recall that word in the list even if they are not supposed to. You can also present a couple of lists like the ones above and tell participants how they are constructed. You can find many lists to use in a paper by Stadler, Roediger, and McDermott (1999). Keep the other instructions the same as in the basic condition, with the warning against guessing. Several experiments have been done using such an instruction, and the general finding is that participants can reduce the level of false recall in the experiment, but they cannot eliminate the effect. It is usually about half as great as in the basic condition with the general warning against guessing. The reduction in experiments measuring false recognition is even smaller (see Gallo, Roediger, & McDermott, 2001).

FIGURE CAPTIONS

Figure 25.1

The DRM false-recognition effect (Roediger & McDermott, 1995, Exp. 2). False recognition of critical lures approximated the hit rate for list items. False alarms to list words from nonstudied lists were 0.11, and those to critical words from nonstudied lists were 0.16.

Figure 25.2

The effects of retention interval on true and false recall (Toglia et al., 1999, Exp. 2). Participants studied five DRM lists, and were given a final free recall test at one of three retention intervals (between-subjects). Data are collapsed across blocked and mixed study presentation, although similar patterns were obtained at each level.