Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 58 (2008) 1065–1079
Journal of Memory and Language
www.elsevier.com/locate/jml

False memories are not surprising: The subjective experience of an associative memory illusion

Jeffrey D. Karpicke [a,*], David P. McCabe [b], Henry L. Roediger III [c]

[a] Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081, USA
[b] Department of psychology, Colorado State University, Fort Collins, CO 80525, USA
[c] Department of psychology, Washington University in St. Louis, St. Louis, MO 63130, USA

Received 24 August 2007; revision received 30 November 2007
Available online 6 February 2008

Abstract

Four experiments examined subjective experience during retrieval in the DRM false memory paradigm [Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22; 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]. Subjects studied lists of related words that were associated with critical non-presented words and then took a recognition test in which they made judgments about their experience of each test item. We tested the prediction from [Whittlesea, B. W. A. (2002). False memory and the discrepancy–attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General, 131*, 96–115] discrepancy–attribution hypothesis that subjects experience critical lures as surprising, and that the experience of surprise leads them to call the lures old. We found that subjects were not surprised when they encountered critical lures on a recognition test and, in fact, they reported that they expected to see critical lures more than they expected to see words that they had actually studied. When subjects did experience words as surprising, they called the words new, not old. The results support the idea that false memories in the DRM paradigm occur when critical lures are activated in memory and fluently processed on a test, leading subjects to experience critical lures in much the same way that they experience words they actually studied. The results do not support the idea that false memories are surprising, as stated by the discrepancy–attribution hypothesis.

Keywords: False memory; Subjective experience; Metamemory; Surprise

One remarkable aspect of false memories is the compelling subjective experience that often accompanies them. For example, in Loftus's misinformation paradigm (Loftus, Miller, & Burns, 1978), when subjects witness an event and then read a narrative that contains misinformation, they falsely remember the misinformation on a later memory test (Roediger, Jacoby, & McDermott, 1996), they attribute the misinformation to the actual event even when told that the narrative contained no true information (Lindsay, 1990), and they are willing to bet money on their false memories (Weingardt, Toland, & Loftus, 1994). In studies of imagination inflation, when subjects repeatedly imagine an event, on a later memory test they will sometimes report

* Corresponding author. Fax: +1 765 496 1264.
E-mail address: karpicke@purdue.edu (J.D. Karpicke).

0749-596X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2007.12.004

that the imagined event actually occurred, even though it did not, and moreover will claim that they consciously remember experiencing the occurrence of the imagined event (Goff & Roediger, 1998). False memories also occur in laboratory tasks with relatively impoverished materials like word lists. When subjects study lists of related words that are associated with a critical non-presented word, they will falsely recall and falsely recognize the critical word at very high levels, and when asked to judge their subjective experience, they claim that they consciously remember experiencing the word when it was presented (Deese, 1959; Roediger & McDermott, 1995; see Gallo, 2006, for review). All of these examples suggest that subjects often experience false memories in much the same way that they experience true memories, and that the compelling subjective experience associated with memory illusions is partly responsible for leading people to mistakenly label their illusory recollections as memories.

One current theory suggests that two sets of processes are involved in the arousal of false memories: Illusory recollections occur when related concepts are activated in memory (Anderson, 1983; Collins & Loftus, 1975), and when monitoring processes that guide decisions about what to call a memory fail to distinguish between events that actually occurred in the past and events that did not occur but were activated in memory (Johnson, Hashtroudi, & Lindsay, 1993). This theory is known as the activation/monitoring framework (Roediger, Balota, & Watson, 2001; Roediger, Watson, McDermott, & Gallo, 2001), similar in many ways to Johnson et al.'s source monitoring framework. The activation/monitoring framework proposes that when subjects study lists of related words that are associated with a non-presented critical word, activation spreads throughout semantic associative networks from the studied words to the critical word, thereby partially activating the critical word. When memory for the words is assessed, failures in the ability to monitor the source of activation during retrieval lead subjects to make the mistaken claim that non-presented critical words were originally studied.

There is considerable support for the activation/monitoring account of false memories in the DRM paradigm (see Roediger et al., 2001; Gallo, 2006). For example, backward associative strength, the degree to which list items tend to evoke the critical item in association norms, is the strongest predictor of false recall (Deese, 1959; Gallo & Roediger, 2002; Roediger et al., 2001). Just as studying semantically related words produces false memories for non-presented semantic associates, studying phonologically related words derived from the same lexical neighborhood (Luce & Pisoni, 1998) also leads to false recall and false recognition of non-presented phonological associates (Sommers & Huff, 2003; Sommers & Lewis, 1999). Further, lists that contain both phonological and semantic associates produce superadditive effects on false recall (Watson, Balota, & Roediger, 2003). Finally, older adults and other populations with deficits in memory monitoring abilities at retrieval show heightened levels of false recall and false recognition (Balota, Cortese, Duchek, Adams, & Roediger, 1999; McCabe & Smith, 2002; Norman & Schacter, 1997). This brief review represents only a small portion of the evidence in favor of the activation/monitoring framework (see Gallo, 2006; for a detailed review), although of course other theories of these phenomena have been developed. The evidence suggests that, in the DRM paradigm, the critical non-presented word becomes activated in memory, and on a test, subjects are unable to distinguish well between non-presented critical words and words that they actually studied. The subjective experiences of critical words and list words seem largely isomorphic.

Another theory explaining illusory recollections is Jacoby's attributional theory (Jacoby, Kelley, & Dywan, 1989), which also holds that subjective experience plays a critical role in guiding people's decisions about what to call a memory, sometimes leading to false memories. The attributional view of memory proposes that people use a fluency heuristic when deciding whether an event is a memory, attributing the fluency of their current processing to indicate that they had experienced an event previously (Jacoby & Dallas, 1981). However, if processing fluency is enhanced by some other means, it may be mistakenly attributed to prior experience. For example, Jacoby and Whitehouse (1989) primed some words in a recognition memory test by briefly flashing the word immediately before it was shown on the test. Priming the words on the test enhanced the fluency of processing those words, leading to increased false alarm rates to non-studied words. (Rajaram, 1993; also showed that this manipulation enhances "know" judgments, in the procedure in which subjects are asked to judge whether they remember or know that they had studied a word). The effects of processing fluency are also observed in other paradigms. Jacoby, Woloshyn, & Kelley (1989) showed that when subjects were asked to judge whether a name was famous, they were more likely to mistakenly judge a non-famous name to be famous when the name had been repeated from an earlier session, because subjects attributed their fluent processing of the repeated name to indicate that the name was famous. Similarly, Jacoby, Allan, Collins, and Larwill (1988) had subjects judge the loudness of a background noise in which words were presented. When words were repeated from an earlier experience, subjects judged the noise to be less loud, even though the objective noise level was the same. Fluent processing of repeated words led subjects to indicate that the noise level was less loud. Just as the activation/monitoring view holds that individuals experience illusory recollections in much the same way as they experience their true memories, the fluency–attribution view also suggests that processing fluency is responsible for both true and false memories, and illusory memories occur when subjects mistakenly attribute fluent processing to the past. Roediger and McDermott (1995) suggested that fluent processing of the non-studied item associatively related to the list might also be partly responsible for the power of the associative memory illusion. Because the critical (non-studied word) is associated to the studied words, it will be processed quite fluently and may give rise to a false sense of familiarity or even recollection.

Recently, Whittlesea and his colleagues (Whittlesea, 2002; Whittlesea & Williams, 1998) have put forth an attributional explanation of false memories, based on the discrepancy–attribution hypothesis, as an alternative to spreading activation theories and Jacoby's fluency theory. According to Whittlesea, Masson, and Hughes (2005), no associative activation needs to be assumed to explain the DRM illusion, and the critical item is not experienced as fluently processed, but instead is experienced as surprising. They stated that:

> [The discrepancy–attribution hypothesis] provides an alternative interpretation of the DRM effect, one that does not involve spreading activation. According to this account, subjective experiences such as the feeling of familiarity are produced by an evaluation process that monitors the integrity of ongoing performance. This evaluation takes into account the apparent quality of current processing as well as those aspects of the current stimulus and context that are salient to the person; it leads the person to develop an attitude toward their performance. Of particular relevance to the DRM effect, this process can sometimes cause the person to experience a perception of discrepancy, a feeling of surprise caused by the apparent mismatch between expectations raised by some parts of a processing experience and the actual outcome. This perception motivates the person to seek an explanation, attributing the surprise to some plausible source in covert characteristics of the current stimulus or environment, the person's current state (mood or disposition), or in the past. When this perception is unconsciously ascribed to a source in the past, the person experiences a conscious feeling of familiarity (p. 421).

The discrepancy–attribution view states that false memories occur when subjects experience a discrepancy between the processing fluency that they expect to experience on a test and the actual fluency of processing that they experience. For example, conceptual factors may lead the subject to expect the critical item (like "sleep" in one of the lists) but the fact that the perceptual form of the item had not been experienced leads to the discrepancy and the conscious feeling of surprise. The critical aspect of the discrepancy–attribution account hinges upon the subjective experience that occurs when individuals have a false memory. Specifically, when a discrepancy occurs between the expected fluency of processing and actual fluency of processing, subjects experience surprise, and this experience of surprise leads them to attribute their surprise to the item being old, thereby producing false memories (see Whittlesea et al., 2005; Whittlesea & Williams, 1998, 2001). As Whittlesea et al. (2005) wrote: "Because they [participants] are unable to understand the cause of variations in their performance, they *consciously experience surprise*" (p. 422, emphasis is ours).

Whittlesea and his colleagues have provided a range of evidence for the discrepancy–attribution account of false memories. For example, Whittlesea and Williams (1998) had subjects study a list that contained words, orthographically irregular nonwords (e.g., "stofwus"), and orthographically regular nonwords (e.g., "hension"). On a subsequent recognition test that included both types of items as lures, subjects showed the greatest levels of false alarms to the regular nonwords. Whittlesea and Williams argued that orthographic processing of the regular nonwords was fluent, because the regular nonwords followed the orthographic patterns found in English, but semantic processing of those nonwords was nonfluent, because the nonwords were meaningless. Further, they argued that the discrepancy between orthographic processing and semantic processing led subjects to experience surprise, and subjects then attributed their surprise to the item being old (but see Cleary, Morris, & Langley, 2007, for an alternative explanation).

In other experiments, Whittlesea and Williams (2001) had subjects study a list of words and then take a recognition test. Each word on the test occurred at the end of a sentence that either predicted or did not predict the word. For example, when the test word was "broom," the sentence, "She cleaned the kitchen with a broom" predicted the occurrence of the target word better than the sentence, "She couldn't find a place to put the broom." In the latter sentence, many more words could potentially complete the sentence stem besides "broom." Whittlesea and Williams also manipulated whether or not a 250 ms pause occurred between the end of the sentence stem and before the target word. They found that false alarms to non-studied words were greater when the words occurred in predictive sentences than when they occurred in non-predictive sentences, but only when there was a pause between the sentence stem and the target word. They argued that predictive sentences led subjects to expect to see the target words that fit the sentence. When a non-studied word appeared after a predictive sentence, with a delay between the sentence stem and the target word, processing of the entire sentence was fluent but the processing of the target word itself was relatively nonfluent. Just as in the "hension" experiment, Whittlesea and Williams proposed that this discrepancy in processing fluency led subjects to experience surprise, and subjects attributed their surprise to the word being old.

Whittlesea (2002) has also argued that the experience of surprise is responsible for false memories in the DRM paradigm (see also Whittlesea et al., 2005). For example, Whittlesea (2002) wrote:

> The illusion of familiarity for prototypes [the DRM associative memory illusion] is not a direct product of their similarity to the study set. Instead, that feeling comes about through two additional steps. First, the similarity causes enhanced production of some, but only some, aspects of processing the prototypes at test. Second, evaluation of this enhanced production leads to the interpretation that it is discrepant with other aspects of processing that are not enhanced... The perception that aspects of a current experience are discrepant occurs when these aspects fit surprisingly well or surprisingly poorly, for a reason that is not immediately clear (hence the surprise)... According to the discrepancy–attribution hypothesis, people experience a feeling of familiarity when certain aspects of their current processing appear surprising, for indefinite reasons, and the past seems to be a plausible source of influence (pp. 97–98).

Thus, according to the discrepancy–attribution hypothesis, when a non-presented critical word like "sleep" appears on a recognition test, the fluency of conceptual or semantic processing is enhanced relative to the fluency of perceptual processing of the critical word, because words related to the critical word were seen in the study phase (enhancing conceptual processing of the critical word) but the critical word itself was never seen (making perceptual processing of the critical word relatively nonfluent). According to Whittlesea and his colleagues, the discrepancy between conceptual and perceptual fluency when processing the critical words causes a conscious experience of surprise, and the experience of surprise is assumed to be the basis of false memories in the DRM paradigm (see Whittlesea et al., 2005; p. 422). Whittlesea has proposed that the discrepancy–attribution hypothesis provides a superior account of false memories in the DRM paradigm because it does not rely on the concept of spreading activation in an associative network (see Whittlesea, 2002; Whittlesea et al., 2005).

Whittlesea et al. (2005) argued that if surprise were the basis for false memories in the DRM paradigm, then a procedure that presumably eliminates surprise should also eliminate the illusion. In their Experiment 3, subjects were shown DRM lists and were instructed to generate the non-presented critical word after study. Whittlesea et al. reasoned that generating the critical word would eliminate the surprise of seeing the word on a subsequent recognition test and, therefore, would eliminate the illusion. Indeed, they found that generating the critical word reduced false recognition. Of course, this procedure does not provide any direct measure of surprise, and there are many other possible explanations for the reduction in false memories that occurred after generating the critical word. For example, Libby and Neisser (2001) also had subjects generate critical words after studying DRM lists and also found reductions in false memories. However, those authors made no reference to "surprise" in their analysis, instead attributing the reduction to a shift from a gist-based strategy to a verbatim-based strategy following generation (cf. Brainerd & Reyna, 2002).

We see several problems with the discrepancy–attribution account of false memories in the DRM paradigm. First, no direct evidence has been produced to indicate that subjects experience surprise when they see non-presented critical words on a memory test, even though the surprise is hypothesized to be conscious (Whittlesea et al., 2005; p. 422). This claim is reasonable, because the concept of "unconscious surprise" makes no sense; the experience of surprise is a conscious one, by definition. Second, a variety of evidence suggests that subjects experience false memories of critical words in much the same way that they experience veridical memories of words that they actually studied. The discrepancy–attribution account, on the contrary, holds that the subjective experience of critical words differs from the experience of other words in that critical words are surprising. Third, the idea that non-presented critical words produce surprise seems inconsistent with other literature on the subjective experience of surprise (for a review, see Reisenzein, 2000). For example, research by Meyer and colleagues, among others, typically induces the experience of surprise by establishing a schema or expectation in the minds of subjects and then violating that expectation (e.g., see Meyer, Niepel, Rudolph, & Schutzwohl, 1991; Meyer, Reisenzein, & Schutzwohl, 1997; Schutzwohl, 1998). In the DRM paradigm, studying lists of associatively related words also establishes a schema by activating associative networks, but non-presented critical words are consistent with the activated networks, and the critical words are falsely recalled and falsely recognized because they are consistent with the network's activation pattern. Based on this prior research on the experience of surprise, it seems plausible that critical words are experienced as unsurprising on the test. However, no prior research has assessed the subjective experience of surprise in the DRM paradigm.

In the present experiments, we investigated whether subjects experienced the occurrence of non-presented critical words in the DRM paradigm as surprising by asking them to make judgments of surprise on a recognition test (Experiment 1). We followed this experiment by asking subjects to judge other types of subjective experience that might, under Whittlesea's theory, be expected to differ for studied words and critical lures, including the expectedness of the words on the test, their unusualness, and their readability (Experiments 2–4). Our goal was to investigate whether subjects experienced studied words and critical lures associated with those words differently, and whether their subjective experience was related to memory performance. The discrepancy–attribution hypothesis suggests that critical lures should be experienced as more surprising than other words on the test, and that this increased surprise should be positively correlated with the likelihood that an item will be judged as "old." In contrast, the activation/monitoring, source–monitoring, and fluency—attribution accounts suggest that studied words and critical lures are both fluently processed and will be given similar surprise ratings. Further, these frameworks predict that words judged to be surprising (and therefore not fitting the associative network activated by the studied lists and not processed fluently) would be called "new."

Experiment 1: How surprised are you to see this word?

In Experiment 1, subjects studied lists of associatively related words and took a recognition test. For each word on the recognition test, they were asked, "How surprised are you to see this word on the test?" They made their responses using a 5-point scale (where 1 = very surprised, and 5 = not very surprised). After making the surprise judgment, subjects then judged whether the word was old or new, using a similar 5-point scale to assess confidence in their recognition decision (1 = sure old, and 5 = sure new). According to the discrepancy–attribution account of false recognition, subjects should experience critical words as more surprising than other words on the recognition test, and the subjective experience of surprise should compel them to call the words old. This relationship between the experience of surprise and the decision to call a word old would be reflected by a positive correlation between the surprise judgments and the old/new judgments.

Methods

Subjects

Twenty Washington University undergraduates, ages 18–22, participated in Experiment 1 in exchange for course credit.

Materials and design

The 28 DRM lists used by Gallo and Roediger (2002) were used in the present experiments. Each list contained 15 words that were associated with a non-presented critical word. Based on the probability of falsely recognizing the critical words in Gallo and Roediger (2002), half of the lists were labeled good lists for producing false recognition, and half were labeled poor lists. For counterbalancing purposes, the 28 lists were divided into 2 sets of 14, with each set containing 7 good lists and 7 poor lists. Half of the subjects studied one set of 14 lists during the study phase, while words from the other set served as distracters on the recognition memory test. The order of sets was reversed for the other half of the subjects. Words from serial positions 2, 7, and 11, as well as the non-presented critical words, served as test items on the recognition test. The recognition test included a total of 112 trials.

Procedure

Subjects were tested in groups of 4 or fewer. They were told that they would study a series of 14 word lists, and that each list would contain 15 words shown simultaneously on the computer screen. The 15 words in each list were shown in two rows of 8 and 7 words, respectively. Although this presentation format is atypical of DRM studies, we aimed to replicate the procedure used by Whittlesea (2002) as closely as possible. Each list was shown for 15 seconds, corresponding to a rate of 1 word per second (see McDermott & Watson, 2001 for discussion of presentation rate and false memory). The 14 lists were presented in a random order for each subject, but the words within each list were always presented in order of forward associative strength to the critical item (Roediger & McDermott, 1995). Subjects were told to read each word one time, silently, and were informed that their memory for the words would be tested later.

After the study phase, the subjects were told that they would take a memory test in which they would see a series of words, some of which were old words that they had studied, and some of which were new words that they had not studied. Subjects made two consecutive judgments about each test word. First, they were asked, "How surprised are you to see this word on the test?" This question was shown on the screen with the test word and a 5-point response scale (labeled 1 = very surprised, 3 = unsure, 5 = not very surprised). After making their judgment of surprise, subjects were asked, "Is this word old or new?" This question was shown on the screen with the test word and a 5-point response scale (labeled 1 = sure old, 3 = unsure, 5 = sure new). After the recognition memory test, the subjects were debriefed and thanked for their participation.

Results and discussion

All results, unless otherwise stated, were significant at the .05 level. Table 1 shows the mean surprise judg-

Table 1
Results of Experiment 1: Mean surprise judgments (1 = very surprising, 5 = not very surprising), probability of calling an item old, and gamma correlations between the surprise judgments and the old/new judgments

Condition	Surprise judgment	Probability old	Gamma correlation
Studied lists			
List items	3.6 (.11)	.57 (.03)	−.70 (.07)
Critical items	3.6 (.13)	.64 (.03)	−.64 (.08)
Non-studied lists			
List items	2.8 (.12)	.18 (.03)	−.64 (.07)
Critical items	2.8 (.14)	.17 (.03)	−.60 (.08)

Note. Standard errors are in parentheses.

ments, the mean probability of calling words old, and the mean gamma correlations between the surprise judgments and the old/new judgments. The probability of calling words old was calculated using old/new judgments of 1 (sure old) and 2 (old). For each subject, and within each item condition, we calculated the Goodman–Kruskal gamma correlation between the 5-point surprise judgment and the 5-point old/new judgment. The results in Table 1 show that levels of false recognition of non-presented critical words were slightly greater than levels of veridical recognition of studied list words. In addition, subjects judged words from non-studied lists as more surprising than words from studied lists, regardless of whether the words were list words or critical words. Subjects did not rate critical words as any more surprising than list words. In all four item conditions, the correlations between the surprise judgments and the old/new judgments were negative, indicating that when subjects did experience a word as surprising, they then identified the word as new, not old, in contrast to the prediction of the discrepancy–attribution hypothesis (Whittlesea, 2002; Whittlesea et al., 2005).

The recognition data (the probability of calling words old) were submitted to a 2 (List type: studied vs. non-studied) × 2 (Item type: critical vs. list items) repeated measures ANOVA. There was a main effect of list type, $F(1,19) = 246.68$, $\eta_p^2 = .93$, indicating that subjects were more likely to identify words from studied lists as old. There was also a marginally significant main effect of item type, $F(1,19) = 2.92$, $\eta_p^2 = .13$, $p = .10$, and a significant list type × item type interaction, $F(1,19) = 3.30$, $\eta_p^2 = .15$. The interaction was driven by the difference in recognition performance for critical words and studied list words: Subjects were more likely to identify critical words as old than they were to correctly recognize list words that they had actually studied (64% vs. 57%, $F(1,19) = 4.59$, $\eta_p^2 = .19$).

The surprise judgment data were also submitted to a 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA. There was a significant main effect of whether or not the lists had been studied, $F(1,19) = 35.35$, $\eta_p^2 = .65$, indicating that subjects judged words from non-studied lists as more surprising than words from studied lists. There was no main effect of item type and no interaction (Fs < 1). Fig. 1 shows the probability of calling a word old (old/new judgments of 1 or 2) as a function of the surprise judgment that the word was given. The figure shows that for both critical words and list words from studied lists, when subjects were surprised, they were not likely to call the words old. The pattern of results shows that subjects treated list words and critical words virtually identically. Subjects were more surprised to see words from non-studied lists than they were to see words from studied lists and, moreover, subjects were not surprised to see non-presented critical lures on the recognition test (relative to list words) even when the list had been presented.

The gamma correlations between the surprise judgments and the old/new judgments were negative in all four item conditions. A 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA revealed no significant effect of list type or item type nor an interaction (all Fs < 1). The relationship between subjects' judgments of surprise and their old/new ratings did not differ across conditions. The consistent pattern of negative correlations indicates that when people did experience words as surprising, they then called those words new rather than old, regardless of whether the words were from studied or non-studied lists and regardless of whether they were list words or critical words.

Experiment 1 showed high levels of false recognition of non-presented critical words, replicating the basic DRM effect. However, subjects were not surprised when they saw non-presented critical words on the recognition

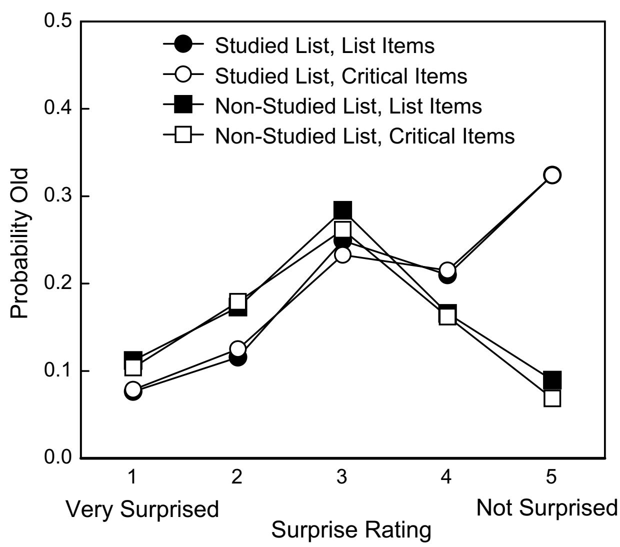

Fig. 1. Probability of calling words old as a function of surprise judgments in Experiment 1.

test, just as they were not surprised to see words that they had studied earlier in the experiment. Furthermore, when subjects did experience surprise upon seeing a word on the test, their experience of surprise was associated with calling the word new, not old, a pattern of results that contradicts the discrepancy–attribution account of false recognition.

Experiment 2: How much did you expect to see this word?

The results of Experiment 1 were opposite those predicted by the discrepancy–attribution account of false memories in the DRM paradigm. The next three experiments were conducted to investigate whether subjects would discriminate between studied words and their associated critical lures based on other types of subjective experience. Although surprise plays a critical role in the discrepancy–attribution account of false memories, other aspects of subjective experience are cited, too. For example, Whittlesea (2002) wrote:

> I therefore suggest that when a subject encounters a prototype [i.e., a critical item following a DRM list] that was prepared by a list of associates, later processing of that item sometimes occurs with much greater fluency than could be expected given the fluency of initial processing... That unexpected fluency of later, semantic aspects of processing is experienced as surprising, giving rise to a powerful perception of discrepancy. In the context of a recognition judgment, this perception of discrepancy is unconsciously attributed to previous experience of the item, causing the person to experience an illusion of familiarity (p. 104).

In Experiment 2, we attempted to explore the subjective experience of surprise again by asking subjects to make a different judgment about their experience during a recognition test. The discrepancy–attribution hypothesis proposes that subjects form an expectation about their processing, and when their actual fluency of processing differs from the processing fluency that they expect, this unexpected discrepancy leads to false memories (Whittlesea, 2002). Thus, in Experiment 2, we assessed subjects' experience of unexpectedness by asking, "How much did you expect to see this word on the test?" using a judgment scale similar to the one used in Experiment 1 (where 1 = not expected, and 5 = very expected). In all other respects the method was identical to that used in Experiment 1. The discrepancy–attribution account of false recognition in the DRM paradigm predicts that people will experience critical words as unexpected, and the experience of unexpectedness will lead them to call the critical words old. This relationship would be reflected by a positive correlation between the expectedness judgments and the old/new judgments.

Methods

Subjects

Twenty Washington University undergraduates, ages 18–22, participated in Experiment 2 in exchange for course credit. None of the subjects had participated in Experiment 1.

Materials, design, and procedure

The materials and design were identical to those used in Experiment 1. The only difference in the procedure was that, for each word on the recognition test, subjects were asked, "How much did you expect to see this word on the test?" The subjects made their judgments of expectedness using a 5-point scale (1 = not expected, 3 = unsure, 5 = very expected), similar to the surprise scale used in Experiment 1. After making their expectedness judgment, subjects were asked, "Is this word old or new?" and made their responses on a 5-point scale (1 = sure old, 3 = unsure, 5 = sure new).

Results and discussion

Table 2 shows the results of Experiment 2. Data from two subjects, who pressed only one response key when making their expectedness judgments, were eliminated from the analysis, and thus the final data set included 18 subjects. The overall pattern of results conceptually replicates the results of Experiment 1. Levels of false recognition of critical words were slightly greater than levels of correct recognition of studied words. In addition, subjects judged words from studied lists as more expected than words from non-studied lists. Most importantly, subjects indicated that they expected to see critical words on the test more than they expected to see words that they had actually studied. The correlations between the expectedness judgments and the old/new judgments were negative in all four item conditions,

Table 2
Results of Experiment 2: Mean expectedness judgments (1 = not expected, 5 = very expected), probability of calling an item old, and gamma correlations between the expectedness judgments and the old/new judgments

Condition	Expectedness judgment	Probability old	Gamma correlation
Studied lists			
List items	3.3 (.12)	.61 (.03)	−.63 (.09)
Critical items	3.6 (.13)	.65 (.03)	−.52 (.12)
Non-studied lists			
List items	2.4 (.12)	.23 (.03)	−.63 (.06)
Critical items	2.4 (.12)	.23 (.03)	−.57 (.08)

Note. Standard errors are in parentheses.

indicating that when subjects experienced a word as unexpected, they then called the word new, not old.

The recognition data were submitted to a 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA. There was a main effect of list type, $F(1,17) = 132.22$, $\eta_p^2 = .89$, indicating that subjects were more likely to identify words from studied lists as old. There was no main effect of item type and no interaction (Fs < 1). As in Experiment 1, subjects were slightly more likely to identify critical words as old than they were to correctly recognize words that they had studied, though this difference was not significant (65% vs. 61%, $F < 1$).

The expectedness judgment data were submitted to a 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA. There was a main effect of list type, $F(1,17) = 60.27$, $\eta_p^2 = .78$, indicating that subjects judged words from non-studied lists as more unexpected than words from studied lists. The effect of item type was not significant, $F(1,17) = 2.54$, n.s., but there was a significant list type × item type interaction, $F(1,17) = 5.28$, $\eta_p^2 = .24$. The interaction was driven by the difference in expectedness judgments for studied list and critical items: Subjects indicated that they expected to see non-presented critical words more than they expected to see words that they had actually studied (3.6 vs. 3.3, $F(1,17) = 5.90$, $\eta_p^2 = .26$).

Fig. 2 shows the probability of calling a word old (old/new judgments of 1 or 2) as a function of the expectedness rating that the word was given. The figure shows a pattern similar to that observed in Experiment 1: When subjects experienced studied words and critical lures as unexpected, they were not likely to call the words old. Fig. 2 shows that subjects experienced studied words and critical lures virtually identically.

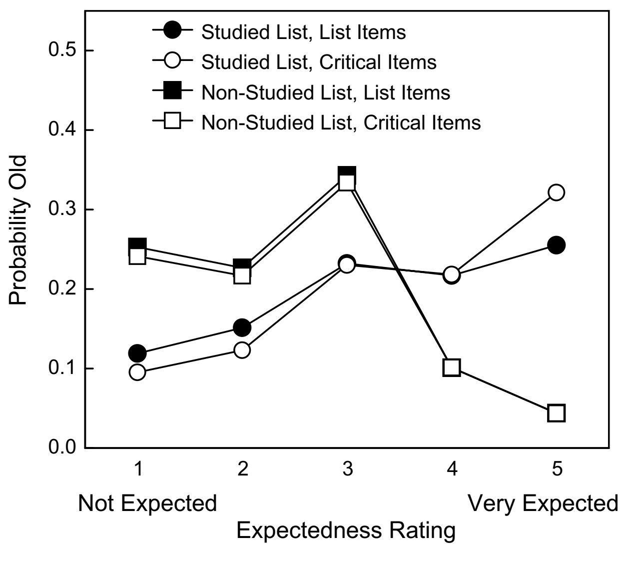

Fig. 2. Probability of calling words old as a function of expectedness judgments in Experiment 2.

The gamma correlations between the expectedness judgments and the old/new judgments were negative in all four item conditions. A 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA revealed a main effect of item type, $F(1,17) = 4.42$, $\eta_p^2 = .21$, but no effect of list type and no interaction (Fs < 1). The pattern of negative correlations in all four item conditions indicates that when people did experience words as unexpected, they then called those words new rather than old, regardless of whether the words were from studied or non-studied lists and regardless of whether they were list words or critical words.

Experiment 2 replicated the results of Experiment 1 and generalized them to judgments of how much subjects expected to see words on a recognition test. Subjects did not experience critical words as unexpected and, in fact, they indicated that they expected to see critical words more than they expected to see words that they had actually studied previously. When subjects did experience a word as unexpected, this unexpectedness was associated with calling the word new, not old, contrary to the prediction of the discrepancy–attribution hypothesis.

Experiment 3: How unusual is this word?

The results from Experiments 1 and 2 showed that subjects do not experience the occurrence of non-presented critical words in the DRM paradigm as surprising or unexpected, and Experiment 2 further showed that subjects expected to see critical words on the test more than they expected to see words that they had actually studied. In Experiment 3, we again attempted to find evidence that subjects experience a discrepancy in their processing of critical words that leads them to falsely recognize the words. Recall that Whittlesea et al. (2005) wrote:

> [Subjects] experience a perception of discrepancy, a feeling of surprise caused by the apparent mismatch between expectations raised by some parts of a processing experience and the actual outcome. This perception motivates the person to seek an explanation, attributing the surprise to some plausible source in covert characteristics of the current stimulus or environment, the person's current state (mood or disposition), or in the past (p. 421).

In Experiment 3, we asked subjects, "How unusual is this word?" to paraphrase the idea expressed in the passage quoted above. Subjects made their judgments of unusualness on a 5-point scale (where 1 = very unusual, and 5 = not unusual). According to the discrepancy–attribution account of false recognition in the DRM paradigm, subjects should experience critical words as more unusual than studied words because of the discrepancy between conceptual and perceptual processing of critical words (conceptual processing of critical words should be fluent, while perceptual processing of the words should be relatively nonfluent; Whittlesea, 2002). Furthermore, experiencing words as unusual should lead subjects to call the words old, and this relationship would be reflected by a positive correlation between judgments of unusualness and old/new judgments.

Methods

Subjects

Twenty Washington University undergraduates, ages 18–22, participated in Experiment 3 in exchange for course credit. None of the subjects had participated in the previous experiments.

Materials, design, and procedure

The materials and design were identical to those used in the previous experiments. The only difference in the procedure was that, for each word on the recognition test, subjects were asked, "How unusual is this word?" The subjects made their judgments of unusualness using a 5-point scale (1 = very unusual, 3 = unsure, 5 = not unusual). After making their unusualness judgment, subjects were asked, "Is this word old or new?" and made their responses on a 5-point scale (1 = sure old, 3 = unsure, 5 = sure new).

Results and discussion

The results of Experiment 3 are shown in Table 3. Once again, the overall pattern of results was similar to the results observed in the previous two experiments. We found robust false recognition of non-presented critical words, although levels of false recognition were not greater than levels of veridical recognition in this experiment. The unusualness judgments showed only small differences across the four item conditions. Furthermore, for non-studied words, the correlations between the unusualness judgments and the old/new judgments were negative, indicating that when subjects experienced a word as unusual, they then called the word new, not old.

The recognition data were submitted to a 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA. There was a main effect of list type, $F(1,19) = 207.64$, $\eta_p^2 = .92$, simply indicating that subjects were more likely to call items from studied lists old. No effect of item type and no interaction were found (Fs < 1). Although false recognition of non-presented critical words was high in Experiment 3, we did not find that critical words were called old more often than studied words, as we did in the previous experiments.

Table 3
Results of Experiment 3: Mean unusualness judgments (1 = very unusual, 5 = not unusual), probability of calling an item old, and gamma correlations between the unusualness judgments and the old/new judgments

Condition	Unusualness judgment	Probability old	Gamma correlation
Studied lists			
List items	3.8 (.15)	.59 (.03)	.00 (.09)
Critical items	3.6 (.18)	.55 (.03)	−.29 (.12)
Non-studied lists			
List items	3.7 (.16)	.20 (.03)	−.38 (.07)
Critical items	3.5 (.17)	.19 (.03)	−.18 (.10)

Note. Standard errors are in parentheses.

The unusualness judgment data were submitted to a 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA. There was a main effect of list type, $F(1,19) = 6.28$, $\eta_p^2 = .25$, indicating that subjects judged words from non-studied lists as more unusual than words from studied lists. There was also a significant main effect of item type, $F(1,19) = 25.26$, $\eta_p^2 = .57$, indicating that, overall, subjects judged critical words as more unusual than list words. Critically though, no interaction was observed ($F < 1$). Thus, although critical words were judged as more unusual than list words overall, this was the case for critical words from both studied and non-studied lists. The small difference in judgments of unusualness appears to be related to some property of the critical words themselves, rather than unusualness of critical words from studied lists, per se (see also Roediger & McDermott, 1995; Whittlesea, 2002). The important point for present purposes is that the discrepancy–attribution framework predicts an interaction between item type and list type, and this was not observed.

Fig. 3 shows the probability of calling a word old (old/new judgments of 1 or 2) as a function of the unusualness rating that the word was given. Overall, when subjects experienced words as unusual, they were not likely to call the words old. As in the previous experiments, Fig. 3 shows that subjects rated their experiences of studied words and critical lures virtually identically.

Finally, for non-studied words, the gamma correlations between the unusualness judgments and the old/new judgments were negative. A 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA did not reveal a main effect of item type ($F < 1$), or a main effect of list type, $F(1,19) = 2.29$, n.s., but there was a significant list type x item type interaction, $F(1,19) = 9.81$, $\eta_p^2 = .34$. This interaction was driven primarily by a difference between studied list items and non-studied list items (.00 vs. −.38; $F(1,19) = 27.13$, $\eta_p^2 = .59$). Studied words showed zero correlation between unusualness judgments and old/new judgments, most likely because subjects were able to base their recognition decisions

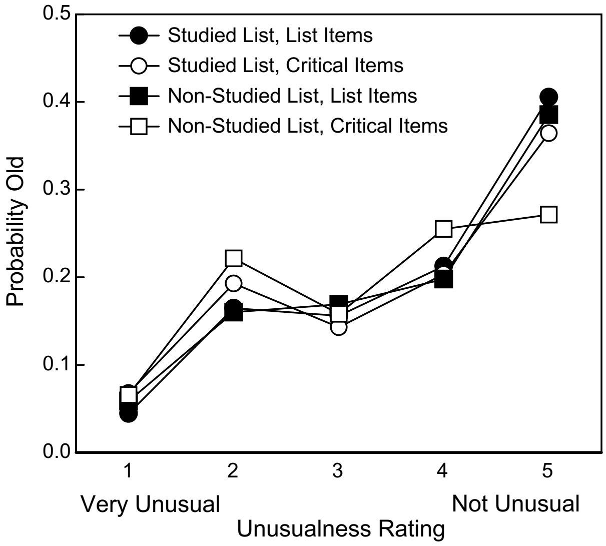

Fig. 3. Probability of calling words old as a function of unusualness judgments in Experiment 3.

on other information (such as recollective detail for these words) rather than how unusual these words appeared to them. However, for the other three types of non-studied items, when subjects experienced these words as unusual, they then called them new rather than old.

Experiment 3 examined the experience of discrepancy by asking subjects to judge how unusual they thought words were during a recognition test. Subjects judged critical words as more unusual than list words overall, but this difference occurred for items from both studied and non-studied lists, indicating that the effect was a result of pre-experimental differences between critical words and list words, rather than a discrepancy in processing produced by studying words associated with critical lures. Most importantly, when subjects experienced a word as unusual, their experience of unusualness was associated with calling the word new, not old, contrary to the idea that the experience of a discrepancy in processing would lead to false recognition.

Experiment 4: How easy was it to read this word?

Experiment 4 investigated a prediction derived from the discrepancy–attribution account. The account proposes that false recognition in the DRM paradigm occurs because critical words are less fluent than subjects expect them to be. Specifically, whereas conceptual fluency of critical words may be enhanced by studying associatively related words, perceptual processing may be relatively less fluent because critical words were never seen in the study phase. To test the idea that critical words are less perceptually fluent than other words, we asked subjects, "How easy was it to read this word on the test?" Subjects made their readability judgments on a 5-point scale (where 1 = easy to read, and 5 = hard to read). The discrepancy–attribution account predicts a negative correlation between readability and recognition judgments: When words are less perceptually fluent, there should be a greater discrepancy between conceptual and perceptual fluency and, therefore, a greater likelihood of judging critical words as "old." On the contrary, the fluency–attribution account (e.g., Jacoby, Kelley et al., 1989; Jacoby, Woloshyn et al., 1989) proposes that critical words are activated at encoding and fluently processed in much the same way as studied words. Thus, subjective fluency of processing studied words and critical lures should be similarly related to recognition judgments, as evidenced by positive correlations between readability judgments and old/new judgments.

Methods

Subjects

Twenty Washington University undergraduates, ages 18–22, participated in Experiment 4 in exchange for course credit. None of the subjects had participated in the previous experiments.

Materials, design, and procedure

The materials and design were identical to those used in the previous experiments. The only difference in the procedure was that, for each word on the recognition test, subjects were asked, "How easy was it to read this word?" The subjects made their readability judgments using a 5-point scale (1 = easy to read, 3 = unsure, 5 = hard to read). After making their readability judgment, subjects were asked, "Is this word old or new?" and made their responses on a 5-point scale (1 = sure old, 3 = unsure, 5 = sure new).

Results and discussion

The results of Experiment 4 are shown in Table 4. Once again, we found robust false recognition of non-presented critical words, and levels of false recognition were slightly greater than levels of correct recognition (as in Experiments 1 and 2). Subjects judged words from studied lists (both list words and critical words) as somewhat easier to read than words from non-studied lists. Importantly, for list and critical words from studied lists, the correlations between the readability judgments and the old/new judgments were positive. Subjects used their judgments of readability when making their recognition decisions, and when they experienced a word as easy to read, they then called the word old.

Table 4
Results of Experiment 4: Mean readability judgments (1 = easy to read, 5 = hard to read), probability of calling an item old, and gamma correlations between the readability judgments and the old/new judgments

Condition	Readability judgment	Probability old	Gamma correlation
Studied lists			
List items	1.8 (.13)	.55 (.03)	.30 (.10)
Critical items	1.9 (.14)	.57 (.03)	.28 (.10)
Non-studied lists			
List items	1.9 (.12)	.21 (.03)	.13 (.08)
Critical items	2.1 (.14)	.22 (.04)	−.04 (.11)

Note. Standard errors are in parentheses.

The recognition data were submitted to a 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA. There was a main effect of list type, $F(1,19) = 102.60$, $\eta_p^2 = .84$, indicating that subjects were more likely to call items from studied lists old. No effect of item type and no interaction were found ($F\text{s} < 1$). As in Experiments 1 and 2, critical words were slightly more likely to be called old than studied list words, though this difference was not significant (57% vs. 55%, $F < 1$).

The readability judgment data were submitted to a 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA. There was a main effect of list type, $F(1,19) = 10.85$, $\eta_p^2 = .36$, indicating that subjects judged words from studied lists as more readable than words from non-studied lists. There was also a significant main effect of item type, $F(1,19) = 7.34$, $\eta_p^2 = .28$, indicating that subjects rated list words as more readable than critical words. However, the interaction was not significant, $F(1,19) = 2.25$, n.s., indicating that critical words were overall rated as more difficult to read.

Fig. 4 shows the probability of calling a word old (old/new judgments of 1 or 2) as a function of the judgments of readability. The figure shows that when subjects judged studied words and critical lures as readable or fluent, they were in fact likely to call the words old. Fig. 4 also shows that subjects experienced studied words and critical lures identically, conceptually replicating the results of Experiments 1–3.

Finally, for list words and critical words from studied lists, the gamma correlations between the readability judgments and the old/new judgments were positive. A 2 (studied vs. non-studied list) × 2 (critical vs. list items) ANOVA on the correlations revealed a main effect of list type, $F(1,19) = 21.71$, $\eta_p^2 = .53$, but no effect of item type, $F(1,19) = 1.06$, n.s., and no interaction, $F(1,19) = 1.38$, n.s. When subjects experienced words from studied lists as easier to read, they then called those words old.

Experiment 4 investigated perceptual processing fluency by asking subjects to judge the readability of words on a recognition test. Critical words were judged slightly more difficult to read than list words but, as in Experiment 3, this difference occurred regardless of whether the items were from studied or non-studied lists, indicating that the readability differences were due to inherent differences in the items, not in the experience of critical words from studied lists. Moreover, when words from studied lists were judged as readable, they were then called old, a result consistent with the idea that critical words are activated in memory during encoding and are processed in the same way as list words, making them subjectively fluent.

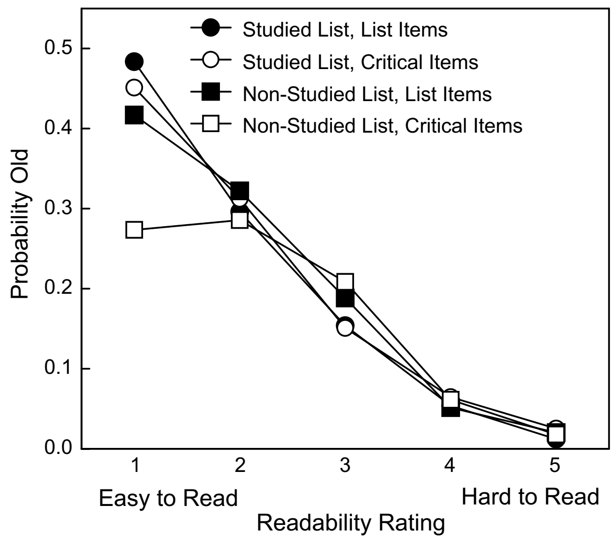

Fig. 4. Probability of calling words old as a function of readability judgments in Experiment 4.

General discussion

In the present experiments, we assessed the relationship between subjective experience and false recognition in the DRM paradigm by having subjects report on their experience before they judged whether test words were old or new. In Experiment 1, subjects were not surprised to see non-presented critical words on the recognition test, and when they did experience surprise upon seeing a word, they then called the word new, not old. In Experiment 2, we replicated the results of Experiment 1 and also showed that subjects expected to see critical words on the test more than they expected to see words that they had actually studied. In Experiment 3, to assess the experience of a discrepancy in processing words on the test, we asked subjects to judge whether each word was unusual. Subjects did not judge critical lures as more unusual than studied words, and when they did experience words as unusual, they then called them new, not old. In short, in the first three experiments we tried to test the subjects' subjective experience using three different measures, each of which accords with hypotheses developed in Whittlesea's discrepancy–attribution theory (Whittlesea, 2002; Whittlesea et al., 2005). All three experiments showed evidence inconsistent with the theory. Finally, Experiment 4 showed that critical words are subjectively fluent, as evidenced by judgments of readability; further, when subjects judged words as readable, they then relied on their overall processing fluency to call the words old, in line with Jacoby, Kelley et al. (1989) and Jacoby, Woloshyn et al. (1989) attributional theory.

The present results are problematic for the discrepancy–attribution account of false recognition, which depends upon the subjective experience of surprise as the proximal cause of false memories in the DRM paradigm (Whittlesea, 2002). We found no evidence that subjects were surprised in processing critical lures in Experiment 1 when we asked them directly about their experience. Similarly, in Experiment 2, when we asked subjects to judge how much they expected to see each word, critical lures were more expected than list items. Overall, subjects did not experience critical words in the DRM paradigm as surprising or unexpected, as would be predicted by the discrepancy–attribution account. We further attempted to assess discrepancy by asking subjects whether they thought critical words were somehow unusual, but the results of that experiment also did not indicate that subjects experienced a discrepancy in their processing of critical words. Finally, the discrepancy–attribution account proposes that false recognition in the DRM paradigm occurs because critical words are less fluent than subjects expect them to be, and a discrepancy between conceptual and perceptual fluency is responsible for false recognition. However, in Experiment 4, subjects indicated that critical words were just as readable as studied words, and this index of perceptual/subjective fluency was positively correlated with the decision to call words old. These results call into question the idea that relatively nonfluent processing of the critical words is partially responsible for false recognition.

Instead, our results are more consistent with the activation/monitoring framework (Roediger et al., 2001), Jacoby's fluency–attribution framework (Jacoby, Kelley et al., 1989; Jacoby, Woloshyn et al., 1989), and Johnson et al.'s (1993) source monitoring framework. Specifically, our results support the idea that critical words are activated in memory after studying lists of related words, and on a memory test, subjects experience critical words in much the same way as studied words, and thus cannot distinguish between the two. The results also are consistent with prior research on the subjective experience of surprise (see Reisenzein, 2000). This research on the experience of surprise employs tasks that establish a schema or expectation in the minds of subjects and then induces surprise by violating that expectation. The DRM paradigm is similar, because studying lists of related words activates an associative network relevant to the list. However, the non-presented critical words are consistent with networks activated by the lists, not discrepant from them; thus, based on other research on the subjective experience of surprise, critical words should not be surprising because they are related to the network activated by the list words. Indeed, our research shows that critical lures are not experienced as surprising, unexpected, or unusual when they follow the relevant list of related words.

We believe that it is precisely because critical words are experienced in the same way as studied words that they are falsely recognized in such a compelling manner. Considerable evidence suggests that subjects experience critical words as indistinguishable from words that they actually studied in terms of both remember/know judgments and confidence ratings (Gallo & Roediger, 2002; McCabe & Smith, 2002; Roediger & McDermott, 1995). However, imagine for a moment that critical words were experienced as surprising, for whatever reason, or that they were simply experienced in some way that differed from other words on the test. If critical words were surprising, why should subjects falsely recognize them? That is, why would subjects not rely on the experience of surprise (or unexpectedness, or discrepancy) to reject the non-presented critical words? Indeed, the activation/monitoring framework suggests that if items are discrepant from activated associative networks they will be more likely to be rejected, but if they are consistent with them they will be endorsed (see McCabe, Presmanes, Robertson, & Smith, 2004). Thus, the present finding that studied words and critical lures were experienced similarly on several dimensions represents further evidence of the subjectively compelling nature of the DRM memory illusion.

A few other problematic aspects of the discrepancy–attribution account of false memories in the DRM paradigm are worth pointing out. First, even if the experience of surprise were responsible for false recognition in the DRM paradigm (and our results show that it is not), it would be difficult to see how the experience of surprise could explain the many other findings in the literature for which the activation/monitoring and source monitoring theories provide a ready account (McDermott & Watson, 2001; McCabe & Smith, 2006; Norman & Schacter, 1997; Roediger & McDermott, 1995; Roediger et al., 2001). According to the activation/monitoring framework, both false recall and false recognition result from similar processes. In both cases there is repeated spreading of activation from studied words to critical lures during the study episode. The summation of this spreading activation results in heightened accessibility of the critical lures in recall, and an increased likelihood of calling the items old, or even "remembered," on a recognition test. Support for this idea comes from studies showing that backward associative strength (i.e., the likelihood of the studied words eliciting the critical lure in word association norms) is the most potent factor that influences recall, recognition, and remembering of critical lures (Deese, 1959; Gallo & Roediger, 2003; Robinson & Roediger, 1997; Roediger et al., 2001). It is not clear how the discrepancy–attribution account could explain the fact that associative strength is the strongest predictor of false recall in the DRM paradigm.

In discussing the discrepancy–attribution account of the DRM false memory effect, Whittlesea (2002) dubbed the effect the ''prototype-familiarity'' illusion. We think this label may hinder conceptual clarity about the nature of the effect for two reasons. First, critical words are associates of list words – indeed, the lists were created by using the first 15 associates of each critical lure – but they are not prototypes in the customary sense used in the categorization literature. Research on categorization has traditionally defined a prototype as consisting of the average of the features of a category (e.g., see Posner & Keele, 1968; Smith & Medin, 1981). By this definition, the critical words in DRM lists are not prototypes of the list words. For example, the critical word ''doctor'' does not represent the average of the features of the words, ''nurse, sick, lawyer, medicine, health,'' and so on. Critical lures are associatively related to list words, but they are not prototypes.. The use of the term ''familiarity'' in the prototype-familiarity illusion label is also unfortunate. Although increases in the familiarity of the critical lure may contribute to the effect, much research has shown that a high proportion of false memories in the DRM paradigm are accompanied by the subjective experience of recollection (Gallo & Roediger, 2003; Geraci & McCabe, 2006; Roediger & McDermott, 2000). In short, calling the DRM illusion the ''prototype-familiarity'' illusion is misleading on both counts.

One might object to our method of assessing surprise directly, suggesting that the attributions subjects make in the DRM paradigm are the result of an unconscious inference. Surprise, however, is by definition a conscious state, and as we noted in the introduction, Whittlesea et al. (2005) have argued that the surprise that gives rise to false memories in the DRM paradigm is experienced consciously:

> In an ideal world, people would understand the causes of variations in their behavior and in each case attribute it to its actual source... However, the participants in such studies are either unaware of or do not understand the effects of such factors. Because they are unable to understand the cause of variations in their performance, they consciously experience surprise. That surprise can, in principle, be attributed to a source in the past, in the covert properties of the stimulus or some characteristic of the person (e.g., mood or skill). Within the context of a remembering experiment, it is likely to be attributed to an unknown source in the past (p. 422).

Thus, our procedure provides a valid assessment of the discrepancy–attribution hypothesis, which boldly proposes that subjects consciously experience surprise when encountering the critical lure on the recognition test. However, subjects report experiencing no such subjective state when processing critical lures and our data show that surprise is negatively, not positively, correlated with false recognition.

Another aspect of our procedure to be addressed is whether asking subjects to rate their surprise before deciding whether words were old or new on a recognition test had any effect on their recognition decisions. We note that in all four experiments we reliably observed robust false recognition of non-presented critical words; in each case false recognition of critical lures was similar to studied items (within 7% in each case), and was very different from false recognition of unrelated items (a difference of 32% or more in each case). This is the typical pattern observed in nearly every experiment using the standard DRM paradigm for studies of recognition memory (Roediger & McDermott, 1995; Experiment 2). Thus, asking subjects to report on their subjective experience before making a recognition judgment did not affect the pattern of results obtained in many prior experiments in which no subjective ratings were obtained prior to recognition. We should also note that although asking subjects to judge how surprising they thought a word was on a recognition test (or how expected, or unusual, or readable it was) was an unusual procedure, subjects were easily able to comprehend and follow the instructions. Furthermore, asking subjects to judge their level of surprise is commonly done in other research directly assessing the subjective experience of surprise (Reisenzein, 2000). Finally, asking for surprise ratings is the most direct test of the discrepancy–attribution hypothesis.

Of course, the subjective experience of surprise may still play a role in other memory phenomena, so our critique is not aimed at the discrepancy–attribution hypothesis per se, but at the application of the theory to the DRM false memory effect. Indeed, surprise may be important to explaining why isolated or distinctive events are remembered better relative to other less distinctive events. Perhaps distinctive items are experienced as surprising when they are encountered, and the experience of surprise has something to do with why they are remembered better later on (e.g., see Hirshman, Whelley, & Palij, 1989; but see also McDaniel, Einstein, DeLosh, May, & Brady, 1995). Of course, distinctiveness effects are produced by making an item different from other items in the same context, and this is opposite from the procedure used to induce false memories in the DRM paradigm, which occurs because critical words are similar to other studied words. Surprise might play a role in the false recognition experiments that the discrepancy–attribution hypothesis was founded on (e.g., Whittlesea & Williams, 1998, 2001). Perhaps words like "hension" are experienced as surprising when they are presented on a recognition test under certain conditions. To our knowledge, the subjective experience of surprise has not been assessed directly in the "hension" experiment or others like it. Nonetheless, the present research using the DRM paradigm points to a straightforward conclusion: False memories in the DRM paradigm are not experienced as surprising, and therefore the discrepancy–attribution hypothesis does not provide a proper account of them.

Acknowledgment

We thank Liz Rhoades for assistance with data collection and scoring.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261–295.

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., & Roediger, H. L. III, (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimers type. *Cognitive Neuropsychology, 16*, 361–384.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*, 164–169.

Cleary, A. M., Morris, A. L., & Langley, M. M. (2007). Recognition memory for novel stimuli: The structrual regularity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 379–393.

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

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

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

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

Geraci, L., & McCabe, D. P. (2006). Examining the basis for illusory recollection: The role of remember/know instructions. *Psychonomic Bulletin & Review, 13*, 466–473.

Goff, L. M., & Roediger, H. L. (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory & Cognition, 26*, 20–33.

Hirshman, E. L., Whelley, M. M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory and Language, 28*, 594–609.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 240–247.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.

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 honor of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General, 118*, 126–135.

Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118*, 115–125.

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

Libby, L. K., & Neisser, U. (2001). Structure and strategy in the associative false memory paradigm. *Memory, 9*, 145–163.

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1077–1083.

Loftus, E. G., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19–31.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1–36.

McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false recognition. *Psychonomic Bulletin & Review, 11*, 1074–1079.

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition, 30*, 1065–1077.

McCabe, D. P., & Smith, A. D. (2006). The distinctiveness heuristic in false recognition and false recall. *Memory, 14*, 570–583.

McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It is not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 422–435.

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.

Meyer, W.-U., Niepel, M., Rudolph, U., & Schutzwohl, A. (1991). An experimental analysis of surprise. *Cognition and Emotion, 5*, 295–311.

Meyer, W.-U., Reisenzein, R., & Schutzwohl, A. (1997). Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion, 21*, 251–274.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838–848.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353–363.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89–102.

Reisenzein, R. (2000). The subjective experience of surprise. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 262–279). Philadelphia: Psychology Press.

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

Roediger, H. L., 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. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association.

Roediger, H. L., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language, 35*, 300–318.

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., & McDermott, K. B. (2000). Distortions of memory. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of memory* (pp. 149–162). Oxford: Oxford University Press.

Roediger, H. L., 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.

Schutzwohl, A. (1998). Surprise and schema strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1182–1199.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Sommers, M. S., & Huff, L. M. (2003). The effects of age and dementia of the Alzheimer's type on phonological false memories. *Psychology and Aging, 18*, 791–806.

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.

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.

Weingardt, K. R., Toland, H. K., & Loftus, E. F. (1994). Reports of suggested memories: Do people truly believe them? In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 3–26). New York: Cambridge University Press.

Whittlesea, B. W. A. (2002). False memory and the discrepancy–attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General, 131*, 96–115.

Whittlesea, B. W. A., Masson, M. E. J., & Hughes, A. D. (2005). False memory following rapidly presented lists: The element of surprise. *Psychological Research, 69*, 420–430.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? The unexpected basis of feelings of familiarity. *Acta Psychologica, 96*, 141–165.

Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy–attribution hypothesis. II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 14–33.