

Running head: REDUCING SOURCE MONITORING ERRORS

Features and Feedback: Enhancing Metamnemonic Knowledge at Retrieval

Reduces Source Monitoring Errors

Sean M. Lane and Cristine C. Roussel

Louisiana State University

Diane Villa

University of Nevada, Las Vegas

Shelby K. Morita

Wells Fargo, N.A.

*In Press – JEP:LMC*

Address Correspondence:

Sean M. Lane  
Department of Psychology  
Louisiana State University  
Baton Rouge, LA 70810  
(225) 578-4098 (office: voice mail)  
(225) 578-4125 (fax)  
E-mail: slane@lsu.edu

## Abstract

Three experiments explored the issue of whether enhanced metamnemonic knowledge at retrieval can improve participants' ability to make difficult source discriminations in the context of the eyewitness suggestibility paradigm. The first experiment documented differences in phenomenal experience between veridical and false memories. Experiment 2 revealed that drawing participants' attention to these differences by pairing the ratings of the features with instructions about their utility were successful in reducing source misattributions of suggested items to the event. The results of Experiment 3 showed that participants can make on-line adjustments in the types of evidence used to make source judgments, as participants who received correct feedback during the training portion of the test reduced misattribution errors on the remainder of the test where feedback was not provided. Altogether, these studies suggest that people can discover and benefit from updated knowledge of the types of memorial evidence that discriminate between sources of information in memory.

*Keywords: source monitoring; retrieval processes; eyewitness suggestibility; feedback*

## **Features and Feedback: Enhancing Metamnemonic Knowledge at Retrieval**

### **Reduces Source Monitoring Errors**

As amply demonstrated by a substantial research literature, people can come to believe they remember perceiving events they actually did not. For example, participants report seeing items in an eyewitness event which had only been suggested afterwards (e.g., Loftus, Miller & Burns, 1978; Zaragoza & Lane, 1994), hearing non-presented words that are highly associated to words which were heard (Deese, 1959; Roediger & McDermott, 1995), and having studied information that was inferred on the basis of relevant knowledge (e.g., Bartlett, 1932; Sulin & Dooling, 1974). Although theories of memory use different postulated mechanisms to account for false memories, most assume that such errors occur because retrieved memory traces are assessed as being similar to veridical memories on at least one characteristic (e.g., Jacoby, Kelley & Dywan, 1989; Johnson, Hashtroudi, & Lindsay, 1993). Thus, for theoretical and practical reasons, there has been much interest in identifying and understanding strategic processes that can be used at retrieval to avoid or edit false memories (e.g., Benjamin & Bawa, 2004; Hicks & Marsh, 1999; Koriat & Goldsmith, 1996; Starns, Lane, Alonzo, & Roussel, 2007).

Although a variety of strategies can be used to avoid memory errors, the present study focuses on the influence of participants' expectations about the retrieval task (metamnemonic knowledge) on their test decisions and how these expectations might be changed in ways that enhance accuracy (e.g., Dodson & Schacter, 2002; Johnson, et al., 1993). In the false memory paradigms discussed above, participants' knowledge about the types of mnemonic evidence that discriminate between veridical and false memories appears incomplete or erroneous, and thus is one potential reason for the high rate of errors observed (e.g., Mather, Henkel, & Johnson, 1997; Schooler, Gerhard, & Loftus, 1986). Identifying whether such knowledge can be updated in

ways that enhance accuracy sheds light on factors that typically promote false memory reports as well as mechanisms which may allow people to learn and adapt their retrieval processes at test.

In most studies utilizing recognition or source monitoring tests, participants are given the explicit goal of accurately discriminating between item classes (e.g., studied vs. unstudied, picture vs. word, event vs. postevent information) and may rely on previous experience with the task, lay theories of memory (e.g., Koriat, Bjork, Sheffer, & Bar, 2004), task information provided by the experimenter, or knowledge acquired during the test (e.g., Dunlosky & Hertzog, 2000) to guide retrieval or post-retrieval decision processes. The extent to which this metamnemonic knowledge accurately reflects task constraints (e.g., the types of items to be distinguished or the type(s) of mnemonic evidence which discriminates between item classes) affects the efficacy of subsequent processes, and ultimately, the accuracy of memory decisions. For example, accurate metamnemonic knowledge could help participants generate retrieval descriptions which effectively discriminate between target and irrelevant traces (e.g., *retrieval orientation*; Herron & Rugg, 2003; see also Jacoby, Shimizu, Daniels, & Rhodes, 2005) or could be used to adopt more diagnostic criteria for evaluating retrieved features of a memory trace (e.g., Johnson et al., 1993). In short, enhanced task knowledge may be translated into a more effective retrieval strategy, thus improving the ability to distinguish between accurate and erroneous memories (e.g., Starns, et al., 2007).

The goal of the following study was to examine the possibility that enhanced metamnemonic knowledge at the time of retrieval can improve participants' ability to make accurate discriminations in the context of a difficult source monitoring situation – the eyewitness suggestibility paradigm (e.g., Lindsay & Johnson, 1989; Loftus, et al., 1978; Zaragoza & Lane, 1994). Participants in studies of eyewitness suggestibility often misattribute postevent

misleading items to the eyewitness event. We sought to improve source monitoring accuracy in two ways. First, we attempted to increase the accuracy of participants' knowledge about specific features which normatively discriminate between memories of event and postevent items by having them rate their memories on these characteristics at the time of test, or by pairing the ratings with instructions about their utility. Second, we sought evidence that participants could discover and update their knowledge about discriminative cues during an initial "training" phase of a source test that provided feedback about the accuracy of their source decisions, and assessed performance on a second phase that lacked feedback. Thus, we examined whether participants could learn from the testing experience to better discriminate between sources and adapt their subsequent processing to improve accuracy.

One impetus for the current study follows from the findings of an extensive literature on the phenomenal experience underlying false memories. This research reveals that one reason false memories are so pernicious is that they can be accompanied by a rich and compelling sense of re-experiencing the context of encoding. For example, participants who falsely claim to have seen misleading postevent suggestions in an eyewitness event often say they vividly remember perceptual and contextual details about its presentation (e.g., Karpel, Hoyer, & Togliola, 2001; Zaragoza & Mitchell, 1996; Schooler, et al., 1986). Although individual false memories can be quite compelling, research has also demonstrated differences in phenomenal experience. For example, specific features of recollective experience (e.g., vividness of perceptual or contextual detail) are typically rated as being lower in erroneous memories than in accurate memories (e.g., Mather, et al., 1997; Norman & Schacter, 1997; Schooler, et al., 1986). Similarly, electrophysiological and imaging studies have found that neural activity consistent with the reactivation of perceptual features occurs more often during accurate than false recognition (the

*sensory reactivation hypothesis*; see Schacter & Slotnick, 2004 for a review). Thus, on average, the recollective experience accompanying false memories appears to be less vivid than for veridical memories. Further, these results suggest that participants may fail to utilize potentially diagnostic mnemonic cues that could improve the accuracy of their decisions.

The above findings raise the question of whether participants are simply failing to notice diagnostic mnemonic cues at test or whether such cues cannot be utilized regardless of their salience. If the former assumption is correct, then participants should be able to utilize these cues if their attention is drawn to them. Prior research suggests that, in some circumstances, highlighting differences in phenomenal experience can improve performance (e.g., Mather, et al., 1997; Schooler, et al., 1986; although see Neuschatz, Payne, Lampinen, & Toggia, 2001). For example, Mather et al. (1997) found that having participants rate their phenomenal experience of recognized items reduced their false recognition of associated non-presented theme words if they had studied intermixed lists, but not if the lists had been blocked by theme. Mather et al. interpreted their findings as suggesting that a more extensive examination of the features underlying a candidate memory trace can be useful in reducing false memories in circumstances where there is not sufficiently strong competing evidence that it was presented (e.g., from strong semantic or schematic activation). Similarly, Schooler et al. (1986) found that participant “judges” were able to distinguish between veridical and false memories generated by other participants, but only when informed of the features which discriminated between such descriptions in prior research (e.g., the presence of sensory details). Although recognition accuracy was improved by highlighting differences in phenomenal experience in these studies, it remains to be seen whether this knowledge would improve accuracy on source monitoring tests. In contrast to old/new recognition, source monitoring instructions specify item classes that

appear on the test and generally lead participants to scrutinize their memories more closely, reducing error (e.g., Dodson & Johnson, 1993; Lindsay & Johnson, 1989). In the following study, we examined whether learning about features that are diagnostic of item class further benefits participants' source monitoring performance.

We conducted three experiments to explore the ability of participants to update metamnemonic knowledge and consequently improve their ability to discriminate between veridical and false memories of a witnessed event. In all experiments, participants saw a mock crime and then answered some questions describing the crime that contained suggestions (objects or actions) which had not been witnessed. Later, participants completed a source memory test. The goal of Experiment 1 was to document differences in phenomenal experience between veridical and false memories of the event. Following each source judgment, participants rated their memories on a number of specific features (e.g., vividness of their memory for object appearance). Participants received the test 10 minutes after completing the postevent questionnaire (Immediate test) or 24 hours afterwards (Delayed test). Based on previous findings, we predicted that veridical memories of event items would be rated as more vivid, particularly for perceptual details, than false memories resulting from postevent suggestion (e.g., Karpel, et al., 2001; Schooler, et al., 1986). In Experiment 2, we tested the hypothesis that participants could use discriminative characteristics identified in Experiment 1 to reduce their misattributions of postevent suggestions to the event. In Experiment 3, we examined the ability of participants to make on-line adjustments in the type of evidence used in their source judgments by varying whether or not they received accurate feedback about their judgments during a pre-assessment "training" phase.

## Experiment 1

### *Method*

*Participants.* A hundred forty-four undergraduate students participated for course credit. There were 84 participants in the Immediate Test condition and 60 in the 24-hour Delayed Test condition.

*Stimuli and materials.* The eyewitness event consisted of a series of slides depicting an office theft (from Zaragoza & Lane, 1994). In the slides, a maintenance man enters an office, repairs a chair, steals a \$20 bill and a calculator, and leaves. There were three versions of the slides that differed only with respect to the presence or absence of six critical items: *a rag, paperback book, jar of Folgers's coffee, Coke can, pack of bubblegum, and a hammer.* Each of these items appeared only in a single slide, and the slide was deleted in alternate versions of the slide sequence. For a given participant, two of these six items were seen only in the slides (slide-only items), two of the items appeared only in the postevent questionnaire (suggested items), and two items served as never-studied control items on the final source test. In addition, six never-seen items (*coat rack, wristwatch, cigarette lighter, newspaper, xerox machine, and wallet*) were used as misleading postevent items. For a given participant, three of these items were suggested only in the postevent questions and three served as control items at test. There were also three items which were seen only in the slides: *typewriter, Lysol can, and Elmer's glue bottle.* On the final test, an additional six items were used as fillers for all participants: five items which appeared in both the slides and the postevent questionnaire and one item that was never presented. Thus, the final 21-item source memory test contained an equal number of items for each of the four possible sources: five slide-only items, five suggested items, five control items, five slide and postevent (Both) items, as well as one never-presented filler item. Items were presented in the same randomized order to all participants.

The postevent questionnaire consisted of 15 paragraphs that each ended with a question. For each participant, five of the questions were misleading in that they suggested information that was not shown in the slide sequence. These items were highly plausible within the context of the event and supplemented rather than contradicted the information presented in the slides. There were six versions of the questionnaire and any specific item served equally often as a suggested and control item. For example, participants for whom the misleading item *coat rack* was suggested were presented with the following question: “At the beginning of the sequence, there was a young woman standing at her desk. As she gathered her purse and blue umbrella from a nearby *coat rack*, what was she preparing to do?” Participants for whom *coat rack* was a control item read a highly similar question in which the word *coat rack* was deleted: “At the beginning of the sequence, there was a young woman standing at her desk. As she gathered her purse and blue umbrella, what was she preparing to do?”

*Procedure.* Participants were tested in groups of up to seven. Upon arrival in the lab, the participants were informed that the experiment concerned people’s interpretation of complex events and that their task would be to watch a series of slides and attempt to understand the depicted incident. The slides were presented at a rate of 4 s per slide. Following the presentation, all participants completed the postevent questionnaire. Participants in the *Immediate* test condition next worked on a word puzzle as a filler task for ten minutes. Participants in the *Delayed* test condition were dismissed and returned 24 hours later.

Finally, participants were given the source memory test in paper and pencil format. Participants were first informed of the various sources of information on the test (see Zaragoza & Lane, 1994 for full instructions). They were also informed that for each test item they would answer two questions: “Saw in the slides?” and “Read in the questions?” Each time they

answered *Yes* to the above questions they had to answer additional questions about the phenomenal experience that accompanied their memory for that item. For items participants claimed to have seen in the slides, they rated the clarity of their memory for the *appearance* of the object in the slides, its *location* in the scene, the *time during the event* when they saw the object, and *thoughts, emotions or reactions* they had when they saw the object on a 1 to 7 Likert scale (1 = very clear memory and 7 = very hazy memory). For those items participants claimed to have read in the questions, they rated the clarity of their memory for *reading the item in the questions*, the item's *location within the paragraphs*, and any *thoughts, emotions or reactions* they had when reading the items on a 1-7 Likert scale. Participants were also asked, yes or no, whether they formed an image of the object when reading the word in the questions, and if yes, rated the ease with which an image came to mind on a 1-7 Likert scale (1 = Very easy, 7 = Very difficult). The test items were read aloud by the experimenter and the presentation was self-paced in that the experimenter waited until all participants had finished answering a particular item. Participants first began with a practice test of three items (not included in the total listed above) to familiarize themselves with the procedure. We note that preliminary analyses which examined critical items (which served as slide-only, suggested and control items across the experiment) and additional slide-only and suggested items separately revealed the same pattern of findings. Thus, in the analyses to follow, we report the mean ratings across all five slide-only and five suggested items.

### *Results & Discussion*

In this and all subsequent experiments, we used an alpha level of .05 for all analyses. The primary dependent measure of interest in this experiment was the ratings given to accurate attributions of slide-only items to the slides (veridical event memories) and misattributions of

suggested items to the slides (false memories). Before discussing the ratings, we first note that over the 24-hour delay, accurate attributions of slide-only items to the slides stayed relatively stable ( $M=.50$  and  $.48$  for Immediate and 24-hour delay tests, respectively), but misattributions of suggested items to the slides appeared to increase ( $M=.29$  and  $.37$ ; for similar findings see Zaragoza & Mitchell, 1996). A  $2 \times 2$  mixed model ANOVA, with time of test as a between-subjects variable and item type (slide-only vs. suggested items) as a within-subjects variable, revealed a significant effect of item type ( $F(1,142) = 34.6$ ,  $MSE = .048$ ,  $\eta_p^2 = .20$ ) and no main effect for time of test ( $F < 1$ ). There was a trend toward an interaction ( $F(1,142) = 3.8$ ,  $MSE = .048$ ,  $p = .055$ ,  $\eta_p^2 = .03$ ).

Given the focus of the series of experiments reported in this article, we report only the rated characteristics of items from the slides which were accurately attributed to the slides and items from the postevent questionnaire which were incorrectly attributed to the slides (i.e., we do not report ratings associated with attributions to the questions). For brevity, means,  $F$  values for comparisons and effect sizes ( $\eta_p^2$ ) are reported in **Table 1**.

*Ratings of Phenomenal Experience.* As is clear from Table 1, the ratings data show that participants generally had more vivid phenomenal experience for the features of accurate than erroneous memories. At the Immediate test, participants rated their memory for the *appearance of the object* in the event as significantly more vivid when the item was actually seen in the slides as when it was read in the postevent questionnaire. Although ratings for the item's *location, time during the event* when the item was seen and *thoughts and reactions* they had when they had seen the item were higher for accurate attributions than erroneous ones, these differences were not statistically significant. Results for the 24-hour delayed test revealed significant differences between accurate and erroneous memories for all rated features except *time during the event*

when the item was seen.

The pattern of results displayed in Table 1 also suggests that the vividness of the phenomenal experience of false memories decreased more over the delay than that of veridical memories. However, there is only modest statistical evidence for this particular point as an omnibus ANOVA (Time of Test x Item Type x Feature) revealed only a trend toward an interaction between Time of Test and Item Type ( $F(1,112) = 3.4$ ,  $MSE = 5.2$ ,  $p = .07$ ,  $\eta_p^2 = .03$ ; the three-way interaction,  $F = 1$ ). Subsequent comparisons revealed that although the average ratings given to slide items did not significantly increase over the delay (higher numbers indicating decreased vividness;  $M = 2.5$  and  $2.8$ , for Immediate and Delayed tests, collapsed across feature type,  $F = 1$ ), there was a significant increase for ratings given to suggested items ( $M = 3.0$  and  $3.7$ ,  $F(1,113) = 4.8$ ,  $MSE = 10.9$ ,  $\eta_p^2 = .04$ ). These findings are consistent with prior research, as Suengas and Johnson (1988) found a proportionally greater reduction in rated vividness over a 24-hour delay for imagined than perceived memories of complex actions.

The results support the claim that, on average, accurate memories are accompanied by a more vivid phenomenal experience than erroneous memories (e.g., Lane & Zaragoza, 1995; Mather, et al., 1997; Schooler, et al., 1986). The largest and most consistent (across time) differences between accurate and erroneous memories were found for the vividness of participants' memory for the *appearance* and *location* of objects. In Experiment 2, we attempted to draw participants' attention to these discriminative mnemonic features as a means of reducing their source misattribution errors.

## Experiment 2

As noted in the introduction, prior research suggests that, under some circumstances, participants may be able to utilize knowledge about differences in the phenomenal experience

which accompanies veridical and false memories to improve performance at the time of test (Mather, et al., 1997; Schooler, et al., 1986). Thus, one would predict that participants in the current study would similarly benefit. However, there are several reasons why such a result may not obtain. First, the postevent suggested items used in this paradigm are highly schema-consistent with the event (i.e., items that would be present in an office), and Mather et al.'s (1997) findings suggest that calling attention to featural information may not be useful in such circumstances (see also Neuschatz, et al., 2001). Second, Schooler et al. (1986) found that instructions to focus on discriminative features were helpful when participants examined written descriptions of other people's memories, but it is far from clear that participants could do so as effectively with the contents of their own memories. For example, working memory capacity is likely much less an issue when looking over a written description than it is when evaluating memory in a series of test decisions.

In Experiment 2, we attempted to draw participants' attention to potentially discriminative features of their phenomenal experience (i.e., object appearance and location) in two ways as a means of reducing their misattributions of suggested items to the eyewitness event. These two conditions differed in terms of the salience of the utility of the provided features for improving the accuracy of source test responses. One group of participants (*Rating-Only* condition) rated their memory of the object's appearance and its location every time they attributed an item to the slides (similar to Mather et al., 1997). A second group was treated identically, but received additional test instructions that informed them of the features' utility and told them to use these features to improve the accuracy of their test decisions (*Instruct+Rate* condition). This is a type of *task validity* instruction because participants receive information about the normative usefulness of the cues for goal achievement (Butler & Winne, 1995). Both

groups were compared to a control condition which received standard source-monitoring test instructions (*Standard* condition). Note that such a control condition is fairly conservative, as source-monitoring instructions and test procedures substantially reduce eyewitness suggestibility relative to old/new recognition (e.g., Lindsay & Johnson, 1989; Zaragoza & Lane, 1994). We predicted that conditions which made diagnostic features salient to participants would reduce false misattributions of postevent suggestions to the slides relative to the control condition, and that the more salient the utility of the features, the greater benefit to accuracy.

## Method

*Participants.* Three hundred forty-one undergraduate students participated for course credit. There were 120 in the Standard instruction control condition, 119 participants in the Instruct+Rate condition, and 102 in the Rating-Only condition<sup>2</sup>.

*Materials and Procedure.* The slides and source memory test items were identical to those used in Experiment 1. The procedure was identical to Experiment 1 until the source test. The source test in this experiment was administered on computer. Participants in the Standard instruction (control) condition received standard source test instructions that described each of potential sources of the items. They were also told that, for each test item, they would answer two questions: "Saw in the slides?" and "Read in the questions?" Participants in the Rating-Only condition were given additional instructions that informed them that every time they claimed to have seen an item in the slides, they would also rate the clarity of their memory for the *appearance* of the object, and its *location* in the event on a 1 to 7 Likert scale (1 = very clear memory and 7 = very hazy memory). Finally, participants in the Instruct+Rate condition received all the foregoing instructions and were also told that previous research had found that when people could clearly recall what the item looked like and where it was located in the scene,

they were more likely to be correct in attributing the item to the slides. Participants were told to use this information to help them discriminate between items they had actually seen and those they had simply read in the questionnaire. The administration of the test items was self-paced. Participants first answered a practice question to ensure they understood the test instructions.

### *Results*

In Experiments 2 & 3, we were primarily interested in the impact of metamnemonic knowledge on participants' source monitoring decisions, particularly for items that were suggested in the context of the postevent questionnaire. Thus, we wished to evaluate whether source memory is enhanced separately from any effects on item memory. Because conditionalized measures are the least likely empirical measures to confound item and source memory (Murnane and Bayen, 1996), we computed the proportion of source responses given to items which were "recognized." Responses to the two questions on the source monitoring test were broken down into four categories (Slide only, Questions Only, Both, and Neither). As in previous eyewitness suggestibility studies (e.g., Zaragoza & Lane, 1994; Zaragoza & Mitchell, 1996), an item was scored as recognized if it was attributed to one or more experimental sources (i.e., "yes" to one or both of the source questions). In the case of suggested items, the conditionalized measure of source accuracy was computed by taking the proportion of items which were attributed *only to the postevent questionnaire* divided by the proportion of items attributed to all experimental sources (i.e.,  $P[\text{Questions Only}] / P[\text{Slide Only}] + P[\text{Questions Only}] + P[\text{Both}]$ ). Source misattribution errors to suggested items involve attributions to the slides (Slide Only + Both responses) and thus the computed proportion was simply the inverse of accuracy. For slide-only items, source accuracy involved attributions to the slides alone, and source misattribution errors involved attributions to the questions alone or to both sources. The

data of primary interest are presented in **Table 2**.

*Source Memory.* We predicted that making diagnostic features increasingly salient to participants would increase their ability to accurately ascertain the true source of postevent suggested items. This prediction was confirmed as there was a significant effect of test condition on participants' attributions of suggested items to the post-questionnaire alone ( $F(2,335) = 6.25$ ,  $MSE = .14$ ,  $\eta_p^2 = .04$ )<sup>3</sup>. As seen in Table 2, as the salience of the manipulation increased, so did accuracy. Subsequent comparisons revealed that Instruct+Rate participants were significantly more accurate than Standard instruction participants, with Rating-Only participants not significantly different from either condition. We also examined the impact of our manipulation on source memory for items from the eyewitness event. Although source accuracy for slide-only items was slightly lower for the Instruct+Rate and Rating-Only conditions than the Standard condition, this difference was not statistically significant ( $F(2,334) = 2.19$ ,  $MSE = .08$ , *ns*,  $\eta_p^2 = .01$ ). Thus, our manipulation increased the accuracy of source memory for suggested items without significantly reducing source memory for items actually viewed in the eyewitness event.

*Misinformation Effect.* Although our primary focus concerned the effect of the manipulation on source memory, researchers studying eyewitness suggestibility are often interested in the overall increase in errors that specifically result from postevent suggestion (*the misinformation effect*; e.g., Loftus, et al., 1978). When examined in terms of source test responses, this effect can be computed as the difference between the unconditionalized misattributions of suggested and control items to the event (Slide-Only + Both responses; Zaragoza & Lane, 1994). An unconditionalized measure of source responding is needed because control items are never presented and thus cannot be accurately "recognized". We report participants' unconditionalized source test responses for all item types in **Appendix A**.

The pattern of data suggests that the manipulation helped decrease errors to suggested items ( $M=.43$ ,  $.33$ , and  $.26$  for Standard, Rating-Only and Instruct+Rate condition, respectively) and control items ( $M = .20$ ,  $.12$  and  $.12$ , respectively). However, there was a significant effect of condition on the size of the misinformation effect ( $M = .23$ ,  $.21$ , and  $.14$ ,  $F(2,338) = 3.6$ ,  $MSE = .08$ ,  $\eta_p^2 = .02$ ). Subsequent tests revealed the Instruct+Rate manipulation significantly reduced the effect relative to the Standard condition. There was also a smaller effect in the Instruct+Rate than the Rating-Only condition, but this difference did not reach conventional significance ( $p < .06$ ). Thus, these findings imply that the provision of instructions and ratings helped reduce the impact of postevent suggestion on participants' test performance.

### *Discussion*

As predicted, calling attention to discriminative features of phenomenal experience increased the accuracy of participants' source judgments about postevent suggested items. Thus, participants were able to correct or update their metamnemonic knowledge of the task and subsequently reduce their misattributions to the eyewitness event. Although participants who simply made ratings of discriminative features benefitted nominally, participants who were explicitly told about the utility of the features showed more substantial gains.

Altogether, the results of Experiment 2 join that of previous work (Mather et al., 1997; Schooler et al., 1986) in suggesting that participants can improve memory accuracy by learning about relevant discriminative features at the time of test. The observed reduction in error is perhaps even more impressive given that the standard source test has been shown to substantially decrease suggestibility relative to standard old/new recognition, and thus participants appear to be already scrutinizing their memories more closely (e.g., Lindsay & Johnson, 1989; Zaragoza & Lane, 1994). This suggests that even with source monitoring tests, participants may be under-

utilizing potentially helpful diagnostic information. Further, our results suggest that participants can benefit from such feature-based instructions even when critical items are highly schema-consistent (cf. Neuschatz, et al., 2001). Finally, we note that these findings are consistent with several experiments from our laboratory using the DRM paradigm (Roediger & McDermott, 1995). Specifically, participants who rated features of their memories and were told to use such features to help them make more accurate judgments reduced their false recognition of critical theme words after hearing studied items in blocked lists (Lane, Roussel, Villa, Starns, & Morita, 2007; cf. Neuschatz, et al., 2001<sup>4</sup>). Thus, this manipulation appears to have utility for improving recognition performance, and more broadly, for other difficult source monitoring situations.

### Experiment 3

Although participants are able to utilize discriminative features of phenomenal experience when told to do so, they often do not appear to spontaneously use such features at the time of test (e.g., Schooler, et al, 1986; results of Exp. 2). In other words, they don't appear to discover such information during the test experience. This suggests that such information is not routinely monitored by participants, or that test conditions don't allow them to detect relatively subtle internal mnemonic cues and associate them with items from different classes. One possible factor is that participants in memory studies don't typically know whether their retrieval decisions are correct or incorrect. Without this information, it may be much more difficult to detect which types of evidence discriminate between veridical and false memories during the test. If this assumption is correct, feedback at the time of test may allow participants to make on-line adjustments in the types of features used to cue memory retrieval or the weight assigned to such features in post-retrieval monitoring (i.e., it may increase calibration). Experiment 3 was designed to test this hypothesis.

Although much has been written on the effectiveness of feedback at test for retention of tested items (for discussion, see e.g., Butler & Winne, 1995; Rosenbaum, Carlson & Gilmore, 2001; Schmidt & Bjork, 1992), the literature on the effect of trial-by-trial feedback on same test performance is substantially smaller. As in the current experiment, these studies have attempted to examine whether participants can use feedback to better calibrate their test decisions. This literature has focused almost exclusively on the effect of feedback on recognition, and the observed impact has been somewhat variable. One of the earliest studies was conducted by Titus (1973); participants who had learned consonant-vowel-consonant trigrams received a recognition test and received feedback or no feedback after each decision. Feedback significantly reduced false alarms and led participants to adopt a more conservative criterion at test. Subsequently, Estes & Maddox (1995) found that feedback at test increased discrimination at marginally significant levels for digit and letter stimuli, but not words. In the most comprehensive set of studies on the topic to date, Lindsay and Kantner (2005; Kantner & Lindsay, 2005; 2006) demonstrated that feedback at test reduced false alarms in recognition with rich, highly structured and temporally extended stimuli (Korean melodies or poetry), but not for stimuli that were relatively simple (e.g., DRM list words, pseudowords). Further, they found that in some situations (e.g., rich, but not temporally extended stimuli such as paintings), participants appear to adopt a more conservative criterion as a result of learning that they will receive feedback (Kantner & Lindsay, 2006). In short, the literature on the effect of test feedback on recognition suggests that type of stimuli plays a major role in determining whether its effect will be increased accuracy, a more conservative response criterion, or null. To our knowledge, there are no published studies that have examined the effect of feedback at test for source monitoring. However, our materials (the eyewitness event) are relatively structured and temporally extended,

suggesting that feedback may be of use in such situations (Kantner & Lindsay, 2005). Finally, an initial pilot study in our laboratory found evidence that feedback at test did indeed reduce source misattributions to postevent suggestions. With this in mind, we conducted a more extensive study to explore its potential effects on source monitoring decisions.

In Experiment 3, we examined whether feedback at the time of test would allow participants to ascertain which features were most discriminative for distinguishing between sources. We did not attempt to assess the nature of these features, but assume they could differ for individuals. As in Experiments 1 & 2, participants saw an event and later received a postevent questionnaire. In addition, we varied the number of times a given item was suggested (once vs. three times). Prior research has demonstrated that repetition increases the confidence and vividness of phenomenal experience associated with source misattributions of postevent suggestions to the event (e.g., Zaragoza & Mitchell, 1996). Thus, the repetition manipulation was used to assess whether feedback varies in its efficacy for reducing errors in situations where participants have strong subjective evidence for its presentation in the event. The feedback manipulation was instantiated by dividing the test into two parts – a training phase and an assessment phase. Participants in our feedback conditions received feedback about the actual source of test items following their responses during the training phase, while those in the no-feedback control condition did not. Our focus was on the assessment phase, where none of the participants received feedback.

In an attempt to discern the effects of feedback at retrieval, we had two feedback conditions. Both conditions received feedback about their source judgments and were told to use it to improve the accuracy of subsequent judgments. In the first condition (*Correct FB*), participants received correct feedback about all the test items in the training phase. In the second

condition (*Incorrect FB*), participants received correct feedback for half of the items and incorrect feedback for the other half. The incorrect items were systematically incorrect in that the provided source was the most common source error for each item type as ascertained by prior research in this paradigm (e.g., never-seen control items are most often misattributed to the event alone; Zaragoza & Lane, 1994). If feedback allows participants to make on-line adjustments in their use or weighting of discriminative features, this would predict that source misattributions of postevent suggestions would be reduced in the Correct FB relative to the Control (No FB) condition. However, such a reduction could also be the result of learning they will receive feedback, which then leads them to become more systematic in their assessment of memory or adopt a more conservative criterion (Kantner & Lindsay, 2006). In this case, both types of feedback should decrease misattribution errors to suggested items relative to the control condition. On the other hand, if participants are relying on the results of the feedback to “tune” their retrieval processing, receiving incorrect feedback should increase errors and receiving correct feedback should reduce them.

### *Method*

*Participants.* A total of one hundred-eighty undergraduate students participated for course credit. The participants were randomly assigned to one of the three experimental conditions (i.e. Correct FB, Incorrect FB and Control), with 60 participants in each.

*Materials.* The eyewitness event was a simulation of a home burglary and a car chase taken from a portion of a training video from the Ohio State Police (see Zaragoza & Mitchell, 1996). Although the event differed from Experiments 1 and 2, the event and materials have been used in numerous studies of eyewitness suggestibility (e.g., Chambers & Zaragoza, 2001; Lane, Mather, Villa, & Morita, 2001) and the pattern of findings have been highly comparable across

the two sets of materials. Further, the materials were developed specifically for studying the impact of repetition of suggestion. The postevent questionnaire consisted of 37 questions regarding the video and was administered on computer. Embedded within these questions was misleading information that was not actually seen in the video. For example, for the suggestion “*the thief wore gloves,*” the participants read and answered the question “At the beginning of the scene, a young man dressed in jeans, a t-shirt and gloves entered the house. Did he enter through the door?” There were 12 critical items (statements) throughout the experiment: *The thief wore gloves, the driver smoked a cigarette, the police thought the driver was DWI, the thief had a gun, the neighbor’s name was Mrs. Anderson, one of the police officers was drinking coffee, the thief took a ring, the thief pulled a window shade down, the driver jumped a curb with the car, the police said they would shoot, the thief put on his seatbelt, and there was a barking dog.* There were three versions of the post–event questionnaire that were equally assigned within each condition, and in any given version four critical items were suggested once, four were suggested thrice, and four did not appear in the questionnaire and served as never-presented control items.

The source memory test consisted of 32 test statements presented on computer. Test decisions were self-paced. The twelve critical items (four suggested once, four suggested thrice, and four control items) were of central interest, as were eight items seen only in the video. There were an additional 12 filler items which were chosen to be clear members of a given source category (eight video + questions items, and four never-presented filler items). The source test was divided in two parts consisting of 16 statements, with the same number of statement types appearing equally often in both parts of the test. The statements were presented on computer in a single random order. Each part occurred equally often across the experiment in the first “training” phase and the second “assessment” phase.

The critical manipulation concerned whether participants received feedback during the training phase of the test (none of the conditions received feedback during the assessment phase). Control participants simply made test decisions and received no feedback. Correct FB participants received accurate feedback after they made a test decision (e.g., “Video Only”). The feedback information remained on the screen for 2 seconds. Incorrect FB participants were treated identically except that half of the test statements were followed by incorrect feedback; for the remaining test statements correct feedback was provided. Whether a given statement was followed by correct or incorrect feedback was counterbalanced across the experiment. The nature of the incorrect feedback varied according to statement type. Specifically, if the correct source was Video Only or Questions Only, the participants were told that the correct answer was “Both in the Video and the Questions.” If the correct source was Both (in the Video and the Questions) or Neither (the Video nor the Questions) source, the participants in the Incorrect Feedback Condition were told that the correct answer was “Video Only.” These choices were based on the most common error types observed in prior research (e.g., Zaragoza & Lane, 1994).

*Procedure.* Participants were run in groups of up to five. They first watched the video and then received the postevent questionnaire. Participants were told that because the video contained so many details, they may have to answer more than one question about some parts of the event. Following the questions, participants worked on a filler task for 10 minutes.

Finally, participants received the source memory test. Test instructions were nearly identical to the previous two experiments, with the exception of additional instructions regarding the format of item presentation. Participants in all conditions were told that they would hear statements that had been *only in the video, only in the questions that they read and answered earlier, both in the video and the questions, or neither in the video nor mentioned in the*

*questions they answered earlier* and that they would have to indicate the source of each test statement (e.g., video, questions, both or neither). Participants were told to indicate their answers (“Video only,” “Questions only,” “Both,” or “Neither”) using the keyboard (corresponding keys were marked V, Q, B or N with stickers). For Part 1 (training phase) of the test, Correct and Incorrect FB condition participants were informed that after making a test decision about a test statement, they would be told the correct source (i.e., they would receive feedback) before the next test statement was presented. Thus, participants would know the accuracy of their source decision immediately. They were further instructed to use this information to assess and improve the accuracy of their judgments. Control condition participants saw the same statements but did not receive feedback. Participants were not told during the first phase that there would be a second phase of the test. Part 2 (assessment phase) of the source test was identical for all conditions. The Control condition was informed that the procedure for Part 2 would be exactly the same as it was for Part 1. The FB conditions were instructed that the procedure for Part 2 would be exactly the same except that they would not receive any feedback as to the accuracy of their judgments.

### *Results*

In Experiment 3, analyses were conducted the same way as for Experiment 2. Although we were primarily concerned with performance during the assessment phase, we first report analyses of training phase performance.

*Source Memory - Training Phase (Part 1).* We analyzed source accuracy for suggested items using a 2 X 3 mixed model ANOVA, with repetition (once vs. thrice suggested) as a within-subjects factor and feedback condition as a between-subjects factor. Although accuracy appeared to improve in participants who received correct feedback, the main effect of condition

was not statistically significant ( $M = .60, .49$  and  $.50$ , for Correct FB, Control, and Incorrect FB condition, respectively;  $F(2,167) = 2.02$ ,  $MSE = .22$ , *ns*,  $\eta_p^2 = .02$ ). Participants were more accurate for once- than thrice-suggested items ( $M = .57$  and  $.48$ ;  $F(2,167) = 4.90$ ,  $MSE = .13$ ,  $\eta_p^2 = .03$ ), but there was no significant interaction ( $F(2,167) = 2.10$ ,  $MSE = .13$ , *ns*,  $\eta_p^2 = .02$ ). We also examined source accuracy for video-only items. Accuracy was somewhat higher in the Control condition than the other groups, but the effect of condition was not statistically reliable ( $M = .88, .93$ , and  $.85$ , for Correct FB, Control, and Incorrect FB, respectively;  $F(2,177) = 2.7$ ,  $MSE = .04$ ,  $p = .07$ ,  $\eta_p^2 = .03$ ). Thus, while feedback was being administered, group differences in source accuracy were small and not statistically significant.

*Source Memory - Assessment phase (Part 2).* We predicted that source accuracy would increase when participants had been given accurate feedback about their source decisions during training, and source accuracy would decrease following incorrect feedback. The pattern of mean performance portrayed in **Table 3** is consistent with this prediction. First and foremost, we found the predicted effect of accurate feedback on source memory for suggested items. A 2 x 3 mixed model ANOVA revealed a significant main effect of feedback condition ( $M = .66, .49$  and  $.43$ , for Correct FB, Control, and Incorrect FB conditions, collapsed across repetition condition;  $F(2,162) = 7.05$ ,  $MSE = .21$ ,  $\eta_p^2 = .08$ ). Subsequent comparisons revealed Correct FB participants were significantly more accurate than the other two conditions, which did not significantly differ from each other. In addition, participants accurately attributed more once- than thrice-suggested items to the questions ( $M = .58$  and  $.47$ ,  $F(1,162) = 5.86$ ,  $MSE = .16$ ,  $\eta_p^2 = .04$ ). There was no significant interaction ( $F < 1$ ). Thus, correct feedback during the training phase was just as effective for increasing source accuracy for items that had been encountered repeatedly during the postevent questionnaire as those which had been read only once. Finally, we examined

source monitoring accuracy for video-only items. There was a significant effect of condition ( $M = .94, .91, \text{ and } .81$  for Correct FB, Control, and Incorrect FB;  $F(2,177) = 7.78, \text{ MSE} = .04, \eta_p^2 = .08$ ), with subsequent tests revealing that incorrect feedback significantly impaired accuracy relative to the Correct FB and Control conditions (which did not differ).

*Misinformation effect.* As in Experiment 2, we examined the impact of our manipulation on the unconditionalized error rate (Video-only + Both responses) to suggested and control items. Unconditionalized responses for all item types are presented in **Appendix B**. The patterns of data show correct feedback reduced errors to suggested items ( $M = .26, .42 \text{ and } .51$ , for Correct FB, Control, and Incorrect FB conditions, collapsed across repetition) and incorrect feedback increased errors to control items ( $M = .06, .08 \text{ and } .16$ , respectively). Thus, there was a significant main effect of condition on the size of the misinformation effect ( $M = .20, .33, \text{ and } .35, F(2,177) = 3.76, \text{ MSE} = .10, \eta_p^2 = .04$ ). The size of the effect was significantly smaller in the Correct FB than in the Control or Incorrect FB conditions, which did not differ. Thus, receiving correct feedback during training allowed participants to reduce the impact of postevent suggestion on test performance.

*Response Time.* Because we recorded the response time of participants at the time of test, we had the opportunity to examine the impact of the manipulation on the speed of participants' judgments. Because correct feedback reduced source misattributions to suggested items, we were most interested in whether these participants took longer to judge the source of suggested items (i.e., engaged in more extensive retrieval processing). We focused on the assessment phase, as all participants received identical instructions and stimuli. We found no significant group differences (collapsed across repetition;  $M = 4464, 4070 \text{ and } 4111$  ms, for Incorrect FB, Control, and Correct FB;  $F(2,177) = 1.9, \text{ MSE} = 1456544, \eta_p^2 = .02$ ), although Incorrect FB

participants were somewhat slower than Correct FB and Control participants. These results suggest that Correct FB participants did not improve their performance simply by examining their memories more extensively. Rather, they appear to have “re-calibrated” the way they cued their memory or weighted facets of their memory during post-retrieval processing.

### *Discussion*

The results of Experiment 3 clearly demonstrate that participants are able to modify their metamnemonic knowledge (although we make no claims about the type of features utilized) and consequently adjust their retrieval processing to more accurately ascertain the source of postevent suggestions after receiving correct feedback about a subset of their judgments. Performance improvement was seen for both once and thrice-suggested items. Further, our results suggest this improved performance was not simply a function of a more conservative criterion or more extended retrieval processes triggered by the feedback instructions. Rather, participants appear to be using the feedback to better calibrate their decisions, as participants who received inaccurate feedback were significantly less accurate than participants who received accurate feedback. Further, Correct FB and Control participants were similar in terms of the speed of source judgments for suggested items. Finally, these results are the first demonstration, to our knowledge, of the utility of trial-by-trial feedback on source test performance. They suggest that it is possible for participants to discover for themselves diagnostic characteristics of item classes and use this information to improve memory accuracy (Odegard & Lampinen, 2006, p. 785).

### General Discussion

Our results can be summarized simply. Participants whose metamnemonic knowledge was enhanced reduced their misattributions of postevent suggestions to the eyewitness event and

increased accurate attributions of these items to their postevent source. We found that this enhancement could occur in two different ways. In Experiment 2, participants showed that they could translate instructions to use potentially diagnostic features of phenomenal experience into more effective source monitoring strategies. Experiment 3 revealed that participants could make beneficial adjustments to their source judgments as a consequence of receiving correct feedback during a training phase. Our findings support the claim that memory errors sometimes occur in false memory research because participants possess incomplete or erroneous metamnemonic knowledge about the nature of the retrieval task (e.g., Mather, et al., 1997; Starns, et al., 2007). More importantly, they demonstrate that it is possible to correct or supplement this knowledge and consequently improve memory accuracy, even under conditions which often produce rich false memories.

Although memory errors can be reduced in a number of ways, our analyses clearly show that the reduction of false memories observed in these experiments was primarily the result of improvements in source monitoring and not due to other strategies such as a general criterion shift. Further, our results suggest that instruction and correct feedback participants particularly benefitted from learning about types of mnemonic evidence (e.g., features of phenomenal experience in Exp. 2) that are diagnostic of source, because their performance improved relative to standard source monitoring participants who were only told about the types of items to be discriminated during the test. According to the source monitoring framework (Johnson, et al., 1993), such information would allow participants to more finely tune the criteria used to evaluate retrieved memories, increasing the accuracy of the resulting decision. Similarly, individuals may use this knowledge to change the cues they use to probe memory during the test (e.g., Herron & Rugg, 2003; Jacoby, et al., 2005), leading to less retrieved “evidence” for the incorrect source.

Regardless, we suggest that more accurate knowledge of the nature of mnemonic evidence which discriminates between item classes allows stronger constraints to be placed on subsequent retrieval and post-retrieval processes, thus improving the ability to ascertain the true source of false memories.

Our data also suggest that participants in eyewitness suggestibility studies may not routinely take advantage of potentially diagnostic mnemonic cues (e.g., features of phenomenal experience) associated with event and postevent items (see also Karpel, et al., 2001). The results of Experiments 2 & 3 demonstrate that this is not due to an inability to use such knowledge to improve performance. In order to detect diagnostic mnemonic cues, participants must monitor for them during the test, and must notice that variations in cues are associated with items of different types (Dunlosky & Hertzog, 2000; Koriat, 1997). Thus, our findings suggest that failure to monitor for these features is a prime reason they are normally underutilized or overlooked (a failure to infer a relationship between specific features and a given source would be another). Further, results of Experiment 3 suggest that the substantial benefits of receiving correct feedback about one's source judgments on later performance may come from an improved ability to monitor and detect such relationships. We intend to examine the nature of the metamnemonic knowledge acquired from feedback in future research.

Although not the focus of this study, our findings naturally invite consideration of whether our manipulations could be used in eyewitness interviews as a means of reducing error. We believe such applications are worthy of exploration, but briefly consider a number of issues regarding their use. Before witnesses could be advised to use specific diagnostic features when evaluating their memories, a key question to be answered is *which* features they should be told to apply. We think it is unlikely the features (object appearance and location) used in our study

would have general utility for distinguishing between genuine memories of eyewitness events and those resulting from postevent sources. Instead, the most diagnostic features for different situations would have to be determined empirically, and then presumably “matched” by an interviewer to fit the given situation (a potentially error-prone process). Feedback could allow a witness to calibrate him- or herself, but there are other practical problems. To provide feedback, an interviewer must first have objective (i.e. corroborative) information about some subset of the case facts. Further, as our findings with incorrect feedback reveal, if the interviewer happened to be wrong they would not only run the risk of suggesting misinformation to the witness, but also impairing accuracy on other event details. Although there may be hurdles to the direct application of our manipulations, the finding that people benefit from enhanced metamnemonic knowledge of task constraints might be fruitfully applied to legal contexts using other techniques. For example, witnesses might receive training to enhance their general ability to evaluate and incorporate task constraints into their memory decisions, or investigators could be trained to help witnesses better assess their own memories during the interview process. Thus, we see reason to be hopeful that research in this area can lead to new ways to reduce real-world false memories.

When faced with a difficult memory task, accurate knowledge of the types of items to be distinguished and the characteristics which discriminate between them can provide important constraints that guide retrieval and post-retrieval processes (e.g., Johnson, et al., 1993; Odergard & Lampinen, 2006; Starns et al., 2007). Conversely, metamnemonic knowledge that is incomplete or erroneous has the potential to negatively impact memory accuracy (Mather et al., 1997). Our findings suggest it is possible to correct or supplement inadequate knowledge by highlighting normatively discriminative features or providing conditions that allow individuals to

discover discriminative features on their own. By helping establish conditions where participants do and do not update metamnemonic knowledge, we see our work as contributing to a broader program that seeks to understand the metacognitive processes that underlie the control of memory performance (e.g., Dunlosky & Hertzog, 2000; Jacoby, et al., 1989; Johnson, et al., 1993; Koriat & Goldsmith, 1996; Reyna & Lloyd, 1997) – an effort with considerable theoretical and practical implications.

## References

- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Benjamin, A. S. & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory & Language, 51*, 159-172.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245-281.
- Chambers, K. L., & Zaragoza, M.S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition, 29*, 1120-1129.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22.
- Dodson, C.S. & Johnson, M.K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology, 106*, 541-557.
- Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46*, 782-803.
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15*, 462-474.
- Estes, W.K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 1075-1095.

- Herron, J. E., & Rugg, M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin & Review*, *10*, 703-710.
- 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.
- 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.
- Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, *12*, 852-857.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28.
- Kantner, J., & Lindsay, D. S. (2005). *Why does feedback improve recognition for unfamiliar melodies but not for words?* Poster presented at the 46<sup>th</sup> annual meeting of the Psychonomic Society, Toronto, Ontario.
- Kantner, J., & Lindsay, D. S. (2006). *When is a false alarm worse than a miss?* Poster presented at the 47<sup>th</sup> annual meeting of the Psychonomic Society, Houston, Texas.
- Karpel, M. E., Hoyer, W. J., & Toglia, M. P. (2001). Accuracy and qualities of real and suggested memories: Non-specific age differences. *Journal of Gerontology: Psychological Sciences*, *56B*, 103-110.
- Koriat, A. (1997). Monitoring one's knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349-370.

- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: the role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*, 643-656.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.
- Lane, S. M., Mather, M., Villa, D., & Morita, S. (2001). How events are reviewed matters: Effects of varied focus on eyewitness suggestibility. *Memory & Cognition, 29*, 940-947.
- Lane, S. M., Roussel, C. C., Villa, D., Starns, J. J. & Alonzo, J. D. (2007). *Providing information about diagnostic features at retrieval reduces false recognition*. Manuscript under review.
- Lane, S. M., & Zaragoza, M. S. (1995). The recollective experience of cross-modality source confusions. *Memory & Cognition, 23*, 607-610.
- Lindsay, D. S., & Johnson, M. K. (1989). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 1077-1083.
- Lindsay, D. S., & Kantner, J. (2005). *Improving recognition memory at test*. Paper presented at the 15th Annual Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science, Montreal, Quebec, Canada.
- Loftus, E. F., Miller, D. G., & Burns, H. G. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19-31.

- Mather, M., Henkel, L., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory and Cognition, 25*, 826-837.
- Murnane, K., & Bayen, U. J. (1996). An evaluation of empirical measures of source identification. *Memory & Cognition, 24*, 417-428.
- Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toggia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory, 9*, 53-71.
- Norman, K. A., & Schacter, D. L. (1997). False recognition in young and older adults: Exploring the characteristics of illusory memories. *Memory and Cognition, 25*, 838-848.
- Odegard, T. N. & Lampinen, J. M. (2006). Memory editing: Knowledge, criteria and alignment. *Memory, 14*, 777-787.
- Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9*, 95-123.
- 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.
- Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual Review of Psychology, 52*, 453-470.
- Schacter, D. L., & Slotnick, S. D. (2004). The cognitive neuroscience of memory distortion. *Neuron, 44*, 149-160.
- Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 171-181.

- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.
- Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal-detection analysis of warning effects in the associative list paradigm. *Journal of Memory & Language, 56*, 592-607.
- Suengas, A.G., & Johnson, M.K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General, 117*, 377-389.
- Sulin, R. A., & Dooling, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology, 103*, 255-262.
- Titus, T. G. (1973). Continuous feedback in recognition memory. *Perceptual and Motor Skills, 37*, 771-776.
- Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 934-945.
- Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science, 7*, 294-300.

### Appendix A

Distribution of Unconditionalized Responses on Source Test as a Function of Test Condition in Experiment 2.

Item Type	Condition	<u>Response</u>			
		Slides Only	Qu. Only	Both	Neither
Slides	Standard	.47 (.02)	.01 (.01)	.06 (.01)	.47 (.02)
	Rating-Only	.47 (.03)	.01 (.01)	.10 (.01)	.42 (.02)
	Instruct+Rate	.45 (.02)	.01 (.01)	.10 (.02)	.44 (.02)
Suggested	Standard	.14 (.02)	.29 (.02)	.29 (.02)	.28 (.02)
	Rating-Only	.11 (.02)	.30 (.03)	.22 (.02)	.37 (.02)
	Instruct+Rate	.10 (.01)	.32 (.02)	.15 (.02)	.42 (.02)
Control	Standard	.16 (.02)	.04 (.01)	.04 (.01)	.76 (.02)
	Rating-Only	.09 (.01)	.04 (.01)	.02 (.01)	.84 (.02)
	Instruct+Rate	.09 (.01)	.04 (.01)	.03 (.01)	.84 (.02)

Note. Standard errors in parentheses. All proportions rounded to nearest hundredth. Attributions are based on joint responses to two separate questions: “Saw in the Slides?” and “Read in the Paragraphs/Questions?” For example, the computed “Slide Only” response involves a ‘Yes’ to the first question and a ‘No’ to the second.

**Appendix B**

Distribution of Unconditionalized Responses on Assessment Phase of Source Test as a Function of Feedback Condition in Experiment 3.

Item Type	Condition	<u>Response</u>			
		Video Only	Qu. Only	Both	Neither
Video	Incorrect FB	.73 (.03)	.03 (.01)	.15 (.03)	.09 (.02)
	Control	.77 (.03)	.00 (.00)	.07 (.02)	.16 (.02)
	Correct FB	.83 (.03)	.00 (.00)	.05 (.02)	.13 (.02)
Suggested X1	Incorrect FB	.18 (.04)	.32 (.04)	.25 (.04)	.26 (.04)
	Control	.13 (.03)	.43 (.05)	.20 (.04)	.24 (.04)
	Correct FB	.12 (.03)	.54 (.05)	.06 (.02)	.28 (.04)
Suggested X3	Incorrect FB	.08 (.03)	.34 (.05)	.51 (.05)	.07 (.03)
	Control	.03 (.02)	.41 (.05)	.47 (.05)	.09 (.03)
	Correct FB	.01 (.01)	.54 (.05)	.33 (.05)	.12 (.03)
Control	Incorrect FB	.12 (.02)	.05 (.02)	.04 (.02)	.79 (.04)
	Control	.08 (.01)	.01 (.01)	.00 (.00)	.91 (.03)
	Correct FB	.06 (.01)	.00 (.00)	.00 (.00)	.94 (.03)

Note. Standard errors in parentheses. All proportions rounded to nearest hundredth.

## Author Note

Sean M. Lane and Cristine C. Roussel, Department of Psychology, Louisiana State University. Diane Villa, Department of Psychology, University of Nevada, Las Vegas. Shelby Morita, Wells Fargo, N.A., San Francisco, CA. The authors thank Adam Botwinik, Courtney Bourgeois, Gerard Broussard, Amy Clark, Will Crescioni, Staci Dahl, Jeff Freehill, David Hays, Stephanie Groft, Melissa Guerrica, Mary Ann Kramer, Dianne Learned, Carla Martinelli, Brent Nobles, Christina Peairs, Stacey Shadoff, Mark Stanga, Todd Thatcher, Kathryn Triche, James Warner, Erin Warnick and Morgan Wolfe for their invaluable help collecting the data, and Zack DeLaune for running a pilot study. We sincerely thank Jason Hicks, Robert Mathews, Emily Elliott, Frank Dempster, and Jeff Starns for helpful comments on earlier version of this manuscript. This research was supported in part by State of Louisiana Board of Regents grant LEQSF(2004-07)-RD-A-12 awarded to S.M. L.

## Footnotes

<sup>1</sup>The degrees of freedom on the ratings analyses differ from the analysis of source attributions because not every participant misattributed a suggested item to the slides. Sixty-four Immediate and fifty 24-Hour Delayed test participants misattributed at least one suggested item to the slides.

<sup>2</sup> Although participants were randomly assigned to conditions, an equipment failure in two sessions led to a slightly smaller N for the Ratings-Only condition.

<sup>3</sup> The degrees of freedom for analyses of conditionalized responses will differ slightly from those on unconditionalized responses when one or more participants fails to recognize at least one item from a given source (e.g., the postevent questionnaire).

<sup>4</sup> In a similar experiment, Neuschatz et al.'s (2001) feature instruction condition reduced critical theme recognition relative to a control group, but the difference was only marginally significant ( $p < .07$ ). Given the strong trend in that study, the results of Experiment 2 and similar findings obtained with the DRM paradigm (Lane, et al., 2007), we believe the evidence is most consistent with the notion that such feature instructions have general utility for reducing false memory.

Table 1. Ratings of Phenomenal Experience for Slide-Only and Suggested Items Attributed to the Slides in Experiment 1.

Measure	Item Source		F (Comparison)	$\eta_p^2$
	Slide-Only (Accurate)	Suggested (Source Error)		
<b>Immediate Test</b>				
Object Appearance	2.4 (.17)	3.1 (.22)	10.1*	.14
Object Location	2.2 (.18)	2.8 (.23)	4.5	.07
Time during Event (when object viewed)	2.5 (.19)	2.8 (.22)	1.9	.03
Thoughts, Emotions or Reactions	3.1 (.23)	3.4 (.24)	1.6	.03
<b>24-hour Delayed Test</b>				
Object Appearance	2.5 (.18)	3.8 (.23)	32.3*	.40
Object Location	2.3 (.19)	3.4 (.26)	22.9*	.32
Time during Event (when object viewed)	2.8 (.23)	3.4 (.25)	5.5	.10
Thoughts, Emotions or Reactions	3.2 (.26)	4.3 (.28)	15.3*	.24

For ratings, lower numbers reflect “clearer” memory for a given type of detail. Standard errors in parentheses. \*  $p < .0125$ ; Bonferroni corrected (.05/4). Degrees of freedom<sup>1</sup> for comparisons: Immediate (1,63) and Delayed test (1,49).

*Table 2. Conditionalized Responses on Source Test in Experiment 2.*

Item Type	Response	Standard	<u>Condition</u>	
			Rating-Only	Instruct+Rate
Suggested	Accurate Attribution	.40 (.03)	.48 (.04)	.57 (.04)
	Misattribution Error	.60 (.03)	.52 (.04)	.43 (.04)
Slide	Accurate Attribution	.86 (.02)	.79 (.03)	.80 (.03)
	Misattribution Error	.14 (.02)	.21 (.03)	.20 (.03)

---

Note. Standard errors in parentheses. All proportions rounded to nearest hundredth. Conditionalized responses are the proportion of recognized items attributed to a source. For suggested items, accurate attributions are to the questions only, and misattributions are attributions to the slides only or both sources. For slide items, accurate attributions are to the slides only, and misattributions are to the questions only or both sources.

*Table 3. Conditionalized Responses on Assessment Phase of Source Test in Experiment 3.*

Item Type	Response	<u>Condition</u>		
		Incorrect FB	No FB	Correct FB
Suggested X1	Accurate Attribution	.45 (.06)	.56 (.06)	.72 (.06)
	Misattribution Error	.55 (.06)	.44 (.06)	.28 (.06)
Suggested X3	Accurate Attribution	.41 (.05)	.42 (.06)	.59 (.05)
	Misattribution Error	.59 (.05)	.58 (.06)	.41 (.05)
Video	Accurate Attribution	.81 (.03)	.91 (.02)	.94 (.02)
	Misattribution Error	.19 (.03)	.09 (.02)	.06 (.02)

---

Note. Standard errors in parentheses. All proportions rounded to nearest hundredth. Conditionalized responses are the proportion of recognized items attributed to a source. For suggested items, accurate attributions are to the questions only, and misattributions are attributions to the video only or both sources. For video items, accurate attributions are to the video only, and misattributions are to the questions only or both sources.