Journal of Experimental Psychology:
Learning, Memory, and Cognition
2001, Vol. 27, No. 2, 362–374

0278-7393/01/$5.00 DOI: 10.1037//0278-7393.27.2.362

The Use of Source Memory to Identify One's Own Episodic Confusion Errors

Steven M. Smith, Deborah R. Tindell, Benton H. Pierce, Todd R. Gilliland, and David R. Gerkens
Texas A&M University

In 4 category cued recall experiments, participants falsely recalled nonlist common members, a *semantic confusion error.* Errors were more likely if critical nonlist words were presented on an incidental task, causing source memory failures called *episodic confusion errors.* Participants could better identify the source of falsely recalled words if they had deeply processed the words on the incidental task. For deep but not shallow processing, participants could reliably include or exclude incidentally shown category members in recall. The illusion that critical items actually appeared on categorized lists was diminished but not eradicated when participants identified episodic confusion errors post hoc among their own recalled responses; participants often believed that critical items had been on both the incidental task and the study list. Improved source monitoring can potentially mitigate episodic (but not semantic) confusion errors.

An eyewitness might falsely recall having seen a red pickup truck among the vehicles in a parking lot if the witness had actually seen the red truck in a different context. Such false memories can be termed *episodic confusion errors,* because events that actually occurred in one episodic context are remembered as having occurred in another episode. In this case, the witness correctly remembers seeing the truck but is mistaken or confused as to the appropriate episodic context of the event. Thus, episodic confusion errors include such phenomena as list discrimination errors (e.g., Anderson & Bower, 1972) and unconscious transference (Wells & Loftus, 1984). Episodic confusion errors constitute one class of false memory. In the present study, we investigated a method for observing episodic confusion errors in recall. We report four experiments that examined circumstances that lead to episodic confusion errors and factors that affect one's ability to identify (and possibly remedy) them.

False memories have been defined as memories of events that did not occur (Roediger & McDermott, 1995). Episodic confusion errors are memories of events that did occur, but they are remembered in the wrong episode. The intrusions commonly observed in false memory studies may be caused more by semantic confusion rather than by episodic confusion. *Semantic confusion errors* are false memories that occur because the incorrect responses are meaningfully related to and confused with events that did occur (e.g., see Brainerd, Wright, Reyna, & Mojardin, 2001). Examples of semantic confusion errors include nonpresented words that are recalled because they are associatively related to studied list words (e.g., Deese, 1959; Roediger & McDermott, 1995), intrusions that occur because they fit meaningfully into schemas (e.g., Brewer & Treyens, 1981), and plausible but incorrect responses that were implied to have occurred (e.g., Loftus, 1991). Some errors might not be caused purely by semantic or episodic confusion but rather by a combination of the two contributing factors. For example, when studying an associatively structured list, participants might momentarily think of the critical nonpresented word, a covert, mental event that later might be confused with the externally presented events. Of the several methods that reliably produce false memories, however, there are none that clearly separate episodic confusion errors from semantic errors. In the associative list method and the schema-guided memory method, the word corresponding to a semantic confusion error is not presented. Thus, the false memories associated with these procedures cannot be clearly traced to an inappropriate event.

The misleading postevent information method might involve episodic confusion errors because experimental participants might confuse suggested material with witnessed material (e.g., Lindsay & Johnson, 1989). If so, however, it may not be possible to unconfound such effects from the effects of inferences made as a result of postevent suggestions. Inferences concerning misleading postevent material, rather than episodic confusion, might be responsible for misinformation effects. Thus, although episodic confusion errors might occur in some studies, the experimental procedures that have been used for observing false memories are not ideally suited for observing episodic confusion errors or for separating out the effects of episodic confusion from the effects of semantic confusion.

The experiments reported in this article derive from a study by Smith, Ward, Tindell, Sifonis, and Wilkenfeld (2000), which showed that common nonpresented category members (e.g., *robin* for *BIRDS*) are often falsely recalled on category-cued-recall tests. Furthermore, Smith et al. showed that such intrusions were more

Steven M. Smith, Deborah R. Tindell, Benton H. Pierce, Todd R. Gilliland, and David R. Gerkens, Department of Psychology, Texas A&M University.

Deborah R. Tindell is now at the Department of Psychology, Wilkes University.

We thank Maria Catala, Todd England, Allison Foster, Holly Frey, Sarah Kragle, Kelli Lingo, Jason Warrington, Steve Wayne, and Scott Wright for their assistance in running the reported experiments.

Correspondence concerning this article should be addressed to Steven M. Smith, Department of Psychology, Texas A&M University, College Station, Texas 77843-4235. Electronic mail may be sent to sms@psyc.tamu.edu.

likely to occur if the critical nonpresented word had been seen in the context of an incidental task. The latter finding (i.e., greater false recall following incidental presentations) is the episodic confusion effect we examined in the present investigation. The episodic confusion effects we report in the present study were always determined in relation to a baseline condition in which no critical items were incidentally presented. Semantic confusion errors were measured by the level of false recall caused by studying semantically related events. Additional false recall can be attributed to recall of incidentally presented items; therefore, the difference between the incidentally presented and nonpresented conditions represents the level of episodic confusion errors.

The episodic confusion effect reported by Smith et al. (2000) would not occur if intruded words were not plausible responses for the category cued recall task. Smith et al. found that only category-relevant intrusions occur with this method. For example, although the words *chair, truck,* and *bluejay* might all be primed on an incidental task, only the semantically appropriate word *truck* was ever given as a (false) response during recall of the vehicles list. Must episodic confusion errors occur as inappropriate responses that exacerbate semantic confusion errors, or can episodically misattributed memories occur in the relative absence of semantic confusion? Some prior evidence indicates that semantic confusion is not necessarily a prerequisite for episodic confusion. For example, another case in which items are remembered from an inappropriate episodic context is a failure of list differentiation (e.g., Anderson & Bower, 1972; Bower, 1972; Winograd, 1968), the ability to determine which of multiple study lists contained a particular target word. List differentiation failures resemble the episodic confusion errors we studied in the present investigation, because both phenomena represent cases in which material presented in one part of an experiment is falsely remembered as having been part of a different to-be-remembered episode. List differentiation failures need not occur because of an exacerbated semantic confusion, per se, although the falsely remembered items may be plausible for other reasons; that is, the incorrect items are words that have the same general characteristics as the correct ones in terms of such factors as meaningfulness, frequency, length, part of speech, and so forth. When recall is guided by preexperimentally acquired semantic knowledge such as scripts, categories, or meaningful associations, there is a greater chance of semantic confusion.

Another related phenomenon, *unconscious transference,* involves misattributing the familiarity associated with previously seen faces (Wells & Loftus, 1984). In unconscious transference, a witness falsely identifies someone (e.g., when identifying the perpetrator of a crime) because that person's face seems familiar to the witness, but the true source of that familiarity is misattributed. In unconscious transference, true events are falsely attributed to inappropriate episodic contexts. This type of episodic confusion does not clearly depend on an underlying semantic confusion, as was the case in our experiments, in which episodic confusion errors were consequences of semantically guided recall.

An episodic confusion error is a type of source-monitoring failure (e.g., Johnson, Hashtroudi, & Lindsay, 1993). Source-monitoring procedures are those that correctly or incorrectly inform the rememberer as to the origin or source of a memory. Source-monitoring failures in memory are cases in which remembered items of information are mistakenly attributed to inappropriate sources. Examples of phenomena attributable to source-monitoring failures include false alarms in an old–new recognition memory task, failures of reality monitoring (e.g., Johnson & Raye, 1981), the false fame effect (e.g., Jacoby, Woloshyn, & Kelly, 1987), and the misinformation effect in eyewitness memory (e.g., Loftus, 1991).

One way to monitor the source of retrieved memories is to recollect material that is episodically bound or associated with the retrieved events. Recollecting the incidental context or the encoding operations used during list learning, for example, may help the rememberer identify the source of an event. In the present investigation, we hypothesize that source information (i.e., information representing aspects of the episodic context of an event) can be used to identify an episodic confusion error as a memory that actually came from a context different than the one in question. The encoding of source information was manipulated in the present study in terms of the level of processing of list items. Deeper processing appears to enhance recollection, as shown by Rajaram (1993), who found that the frequency of *remember* judgments in recognition, an indicator of recollection, increased with deeper processing. Therefore, it was predicted that deeper processing during study would endow memories with more potential source information, enabling participants to better avoid and identify their own episodic confusion errors on a recall test, even though the deeper level of processing might make it more likely that the person would generate the word corresponding to the episodic confusion error at the time of retrieval.

Some types of false memories have proven difficult for people to avoid. For example, a popularly observed semantic confusion error is false recall of the word *spider* when associates to the word, such as *arachnid* and *web,* have been presented (e.g., Roediger & McDermott, 1995). McDermott and Roediger (1998) found that participants had difficulty avoiding such semantic confusion errors when they recalled associatively structured lists of words. In spite of warnings they heard about intrusions that were associated with each of the list words, participants were unable to avoid recalling the critical nonpresented linking words (but see Gallo, Roberts, & Seamon, 1997).[1] Apparently, those false memories of nonpresented words could not be easily distinguished from veridical memories of presented words.

McDermott and Roediger (1998) reasoned that according to the source-monitoring framework, people should be able to distinguish presented words from nonpresented ones on the basis of the perceptual information that can be retrieved in the service of recollection. That is, more associated perceptual detail (i.e., source information) should be accessible for presented words than for nonpresented ones. If such is the case, stated McDermott and Roediger, then their participants often did not take advantage of such source information. Hicks and Marsh (1999) demonstrated a method for augmenting or encouraging the encoding of specific

[1] Gallo et al. (1997) reduced false recognition of critical items from associative lists when participants were forewarned, before studying the word list, about the nature of the critical linking words. Such a reduction in the false recognition effect does not appear to be based on correctable source monitoring. Rather, a combination of deducing the identity of the linking word during study and monitoring for that word during list presentation could reduce the false recognition effect.

source material, such as speaker's voice or anagram solving operations, which reduces false recall of critical nonpresented items in the associative list method. These results offer promising directions in terms of reducing or eliminating false memories.

Having accurately encoded source memory for an event does not by any means guarantee accurate source monitoring when the event is remembered. An example of participants not bringing source memory to bear on a memory task was demonstrated in a cryptomnesia study by Marsh, Landau, and Hicks (1997). This study found that source memory was poor when participants generated ideas that should have excluded responses given in an earlier session (referred to as a *generate-new* task). Nonetheless, source memory for the ideas was shown to remain intact on a different *identify-other* task, in which participants identified their own plagiarisms from the earlier session. Although Marsh et al.'s participants did not use source memory on the generate-new ideas task, the fact that source memory had been stored was shown in the identify-other ideas task. Further evidence of the limitations of using source-monitoring tasks to reduce false memories was provided by Hicks and Marsh (2001), who found that, relative to a standard recognition test, a source-monitoring test actually produced greater false memories.

The possibility that McDermott and Roediger's (1998) participants had untapped source memory that could have been (but was not) used to differentiate intrusions from correctly recalled words seems doubtful. Their participants were given explicit instructions about the nature of the linking words and were adequately forewarned about the errors. If McDermott and Roediger's participants had such source knowledge, it seems likely that they would have used it. It seems that increasing one's source-monitoring efforts cannot improve one's ability to reject semantic confusion errors, false memories that are not associated with episodic perceptual detail (but see Mather, Henkel, & Johnson, 1997).

Although the presence of appropriate source knowledge during recollection may be diagnostic of an item's list membership (Mulligan & Hirshman, 1997), the lack of appropriate source information may not prove that an item did not belong to a memorized list. For example, if study conditions are such that many target list words are encoded without much source information, then impoverished source information during recall might not distinguish a false memory from an accurate one. Thus, semantic confusion errors produced during recall, which lack appropriate source knowledge, are indistinguishable (to the rememberer) from semantically and episodically correct responses that also lack source knowledge (e.g., because of poor binding of contextual associations). In contrast, in the present study, episodic confusion errors could be identified by the presence of specific source information rather than by the absence of source information, as in McDermott and Roediger's (1998) study. If primed items are retrieved that are members of the category being recalled, their inappropriate source information (associations with the priming task) can be used to identify them as errors. In the present study, the source information for episodically inappropriate events could be used for identifying such intrusions as events from an incidental word rating or vowel counting task. The source information associated with episodic confusion errors is different from that of correct memories, making such source information diagnostic in terms of rejecting episodically inappropriate memories during or after recall.

Episodic confusion errors in recall might occur when automatic retrieval of incidental items is unopposed by accurate recollection of the source of the incidentally presented word. The method used in our experiments allows the observation of "opposition" of memory processes, in which the automatic retrieval of an incidentally presented item should be opposed by conscious recollection of the item (e.g., Jacoby, 1991). An example of opposition is the false fame effect, the finding that nonfamous names that had been studied are falsely remembered by participants as being famous names (e.g., Jacoby et al., 1987). Another example of a method that uses this type of opposition is a stem completion or anagram solution task, in which previously encountered words are forbidden as solutions (e.g., Jacoby et al., 1987). Although automatic and intentional memory processes might typically function together to produce appropriate responses in many episodic memory tasks, in such opposition paradigms the two processes may function to produce inappropriate responses. Inappropriate responses are given, according to this explanation, when automatically retrieved material is not attributed to its correct episodic source, a failure of conscious recollection. The episodic confusion phenomenon we investigated can likewise be explained; automatic retrieval of incidentally presented items, combined with failure to accurately recollect the episodic source of those incidentally presented words, may produce a portion of the false recall observed with this task.

The general procedure for all four experiments is shown in Table 1. In each experiment there was an incidental task containing critical primes for half of the categorized lists, plus many other unrelated words. The incidental task involved shallow or deep processing of the words, a between-subjects variable. Participants then studied eight categorized lists of words, followed by eight category-cued-recall tests. In Experiments 2 and 3, after the last cued-recall test, we gave a source recognition test in which participants reviewed all of their cued-recall responses and identified the source of each response (e.g., seen in the incidental task or on the categorized list). In Experiments 1 and 4, source-monitoring instructions were given prior to the cued-recall tests. At the end of each procedure, a final free-recall test was given, requesting recall of the words that had been seen on the incidental task at the beginning of the experiment. This final free-recall test was used as a manipulation check on the effect of levels of processing on recall of primed words, the results of which are shown in Table 2.

The four experiments in the present study examine various aspects of episodic confusion errors. Experiment 1 tested participants' ability to include and exclude episodically confusable words in category-cued recall, as a function of the input level of processing of the incidentally presented critical word. We predicted that deeper processing would endow items with more source information, thereby allowing more accurate recollection of source. Therefore, deeper processing of incidentally presented critical words should lead to an improved ability to avoid episodically inappropriate items during recall or to include incidentally presented items in cued recall if instructed to do so. Experiments 2 and 3 tested the ability to identify episodic confusion errors after the fact. That is, we tested whether or not participants could identify which of their recalled words had actually been presented on an incidental task.

We also predicted in Experiment 3 that on a task involving source recognition of one's own episodic confusion errors (i.e., recognition that such words were presented on the incidental task), levels of processing effects should be restricted to "remember"

Table 1
General Procedure for Experiments 1–4

1. Incidental task
16 unrelated words including four *critical items* (e.g., cloud, glass, *bluejay*, mountain)
Level of processing: deep or shallow (between-subjects)
Deep: Rate how each word makes you feel
Shallow: Count the vowels in each word
2. Categorized list learning
Study eight categorized lists, each missing a high output dominance critical item
Four lists correspond to critical items shown on incidental task, four others not presented
3. Category cued recall
Given category name (e.g., BIRDS, FURNITURE), recall words presented on study list
Experiment 1: Include or exclude or standard recall instructions
Experiments 2 and 3: Standard recall instructions
Experiment 4: Recall with source monitoring instructions
4. Source recognition test (Experiments 2 and 3 only)
Identify the source (e.g., incidental task, study list) of recalled words
Experiment 3: Remember–know judgments
5. Final free recall
Free recall of the 16 unrelated words from the incidental task

responses and are not expected to affect "know" responses. Deeper processing of critical words during the incidental task should increase the likelihood that when such words are falsely recalled, participants will claim to recollect having seen those words on the incidental task ("remember" judgments). Deeper processing should not affect the number of falsely recalled critical items that participants simply know were on the incidental task ("know" judgments).

Deeply processed source information associated with a critical item should make the item identifiable as an event from the incidental priming task. Do participants have the illusion that critical items were presented twice: once on the incidental task, and again on the categorized list? Experiment 4 again tested whether the episodically confused words could be attributed to the incidental task and also whether those items could be correctly identified as not having been members of the studied categorized lists. Does source information associated with words processed deeply on the incidental task eliminate the illusion that the critical words also appeared on the categorized lists?

We expected false recall of the critical words to occur in all experiments. That is, in all conditions, the most common category member omitted from each categorized list was expected to be falsely recalled, as in Smith et al.'s (2000) experiments. The level of false recall when the critical words were not presented in an incidental task, referred to as a *semantic confusion effect,* was used as a baseline from which to calculate the *episodic confusion effect.* The magnitude of the episodic confusion effect was calculated as this baseline level of false recall of critical items subtracted from the level of false recall when critical items are included on the incidental task (e.g., Smith et al., 2000). We did not expect the level of correct cued recall to be affected by any of the experimental manipulations; the results of this measure for all four experiments are shown in Table 2.

Table 2
Mean Proportions Correct (and Standard Errors) for Cued Recall and Final Free Recall as a Function of Level of Processing in Experiments 1–4

	Level of processing			
	Shallow		Deep	
Experiment	*M*	*n*	*M*	*n*
1				
Correct cued recall	.447 (.009)	119	.437 (.009)	114
Final free recall	.228 (.011)		.572 (.019)	
2				
Correct cued recall	.539 (.039)	50	.494 (.019)	49
Final free recall	.163 (.019)		.657 (.025)	
3				
Correct cued recall	.464 (.016)	50	.441 (.014)	59
Final free recall	.181 (.019)		.558 (.028)	
4				
Correct cued recall	.439 (.015)	37	.449 (.015)	48
Final free recall	.264 (.029)		.602 (.032)	

General Method

Participants

The participants in all of the experiments were undergraduate volunteers who completed part of a course requirement by their participation. Participants were recruited for group sessions using posted sign-up sheets. Volunteers enrolled for any one of many experiments, including ours. There were unequal numbers in the treatment groups, because unequal numbers of participants enrolled for different experimental sessions. Each session was held in a group of approximately 5–15 participants at a time. There were 234 volunteers who participated in Experiment 1, 99 in Experiment 2, 109 in Experiment 3, and 92 in Experiment 4. The number of participants in each experimental condition is shown in Table 2.

Materials

We drew eight categorized lists from Smith et al.'s (2000) materials. Each list corresponded to a different category, such as *clothing* or *vehicles* (Smith et al., 2000, Appendix A). Each list contained 15 items from a single taxonomic category. The category member highest in output dominance (Barsalou, 1985), referred to as the critical item, was omitted from each list, and the next 15 most dominant exemplars were included on each list.

Sixteen unrelated common English nouns were presented on the incidentally presented list of words. The incidental list consisted of 12 filler nouns that were not members of any of the categories used in the experiment, plus 4 critical words, each with a nonpresented word corresponding to one of the eight subsequently presented categorized lists. The 4 critical items in one counterbalancing corresponded to the odd-numbered (in terms of presentation order) categorized lists, and the 4 in the other counterbalancing corresponded to the even-numbered categories.

Stimuli were recorded on videotape from the output of a computer program and were shown to participants on a large television screen. Each word was shown individually on the screen, using large black uppercase letters on a white background.

Design and Procedure

The experimental session began with a task in which participants rated a list of unrelated words. This task is referred to as the *incidental task,* which was used as a means for presenting critical items that were not on the subsequent categorized lists. The purpose of this word-rating task was to present half of the critical items in an incidental task, inducing either deep or shallow input processing of the critical words. The variable *incidental presentation* refers to whether or not a critical item was presented on the incidental task and was a within-subject factor. Level of processing was manipulated between-subjects; participants either rated words on a pleasantness scale (deep processing) or counted the number of vowels that were contained in each word (shallow processing). Each of the 16 incidental list words was shown on a television screen for 5 s, during which time participants either counted the vowels in the word or rated the word's pleasantness on a scale of −3 (*very bad*) to +3 (*very good*). Participants recorded responses to each trial on a paper response form. After all of the words had been seen in the incidental task, the response forms were collected.

The next task, the presentation and study of the eight categorized lists, was presented as a new task unrelated to the first. Participants were instructed to view and study the eight categorized lists to prepare for a subsequent unspecified type of memory test over the lists. Recall tests were delayed until all eight lists had been studied. At the beginning of each list presentation, the category name (e.g., *birds* or *furniture*) was shown for 3 s, after which the 15 list items were shown. Each of the 15 items on a list was shown on the television screen for 1.5 s.

Category names were subsequently used as recall cues for each list. On each cued recall test, a category name appeared on the screen. Each recall test lasted 2.5 min. The eight cued-recall tests were given in the same order as the studied lists. A 10-s pause was given after each cued-recall test.

The recall instructions for the cued-recall tests in Experiment 1 were either standard, inclusion, or exclusion instructions. The type of instructions given was a between-subjects variable. The standard instruction, which was used in all conditions for the cued recall tests in Experiments 2 and 3, specified that participants should recall only words from the studied categorized lists. In Experiment 1, exclusion instructions stated that some of the words from the incidental task belonged to the same categories as the studied lists and that such items *should not* be written down on the cued-recall tests. Inclusion instructions stated that words from the incidental task *should* be recalled with their appropriate categories on the cued-recall tests.

In Experiment 4, all participants were given the same cued-recall instructions. These instructions stated that participants should write down not only the category members that had appeared on the study list but also any incidentally presented items that were from the same category. Furthermore, for each word recalled, participants wrote a *1* in the margin to indicate that the word came from only the incidental list, a *2* if the word had appeared only on the studied list, and both a *1* and a *2* if the word had appeared on both lists. Participants wrote down these numbers online, as each word was being recalled.

In Experiments 2 and 3, after the eight cued-recall tests were completed, participants were asked to look over all of the words they had recalled for each categorized list and to circle any of their recalled words that they had seen in their original incidental task (described to participants as the pleasantness rating task and vowel counting task). This task is referred to as the *source recognition test,* because participants were asked to recognize the true source of their recalled responses. The source recognition test was self-paced. As in all other experiments, level of processing was manipulated between-subjects, and incidental presentation was manipulated within-subjects. Participants in Experiment 3 were additionally instructed to indicate whether they remembered having seen or thought about the recognized items on the word rating task or whether they simply knew the items had been on the initial task. A "remember" judgment on this test was defined for participants as an item they remembered seeing or thinking about when it appeared on the incidental task. A "know" judgment was defined as one in which participants knew the item was on the incidental task but could not remember seeing or thinking about that item on the incidental task. Participants appeared to understand the remember–know distinction quite readily.

Finally, in all four experiments, participants were asked to recall the words they had seen on the original incidental (vowel counting or pleasantness rating) task. This final free-recall test for the incidentally presented words was used as a manipulation check to determine whether the level-of-processing manipulation had the desired effect, with shallower processing leading to poorer recall (see Table 2).

Experiment 1

Experiment 1 tested whether participants were capable of avoiding episodic confusion errors by bringing source memory to bear during the cued-recall tests. Before the category-cued-recall tests began in Experiment 1, we told participants that some of the words from the incidental task were members of the categories that were subsequently studied. In the exclude condition of Experiment 1, participants were instructed not to write down any such incidentally presented items on the cued-recall tests, whereas participants in the include condition were told to write down any incidentally presented category relevant items on the cued-recall tests. Jacoby (e.g., 1991, 1998) used this "method of opposition," placing recollection and automatic processes at odds, in terms of the task requirements, to estimate the relative contributions of these two memory processes in task performance. Participants in the standard instruction condition in Experiment 1 of the present study were told to recall only words that had been presented on the categorized lists. If source memory is stored but not typically used on the category-cued-recall test, then participants in the exclude condition should write down fewer incidentally presented critical items on the cued-recall tests, and those in the include condition should give incidentally presented items as responses more often on the cued recall tests, relative to the standard instruction condition. Therefore, we predicted that participants who deeply processed incidental items would be able to reliably include and exclude such items on the cued-recall tests. If they could reliably exclude critical words, participants should be able to reduce false recall of critical words to the level of the baseline condition, in which the critical word was not incidentally presented. To successfully include incidental words in recall, participants should recall critical items reliably more often than in the false recall baseline. We predicted that those who shallowly processed incidentally presented items would not be able to reliably include or exclude those items during category-cued-recall, meaning that recall of presented critical words should not reliably differ from the baseline.

Results

A significance level of $p < .05$ was used on all statistical tests for all experiments, unless otherwise specified. Level of processing was a between-subjects variable with two levels, shallow and deep, in all four experiments. Incidental presentation was a repeated measure with two levels, incidentally presented (i.e., primed) or not presented (unprimed), in all four experiments.

Correct Cued Recall and Final Free Recall for all Experiments

Level of processing of the incidental list had no noticeable impact on correct cued recall of the categorized lists in any of the four experiments: Experiment 1, $F(1, 231) = 1.00$; Experiment 2, $F(1, 97) = 0.23$; Experiment 3, $F(1, 107) = 1.12$; Experiment 4, $F(1, 83) = 0.21$. We computed another one-way analysis of variance (ANOVA) comparing recall of the 16 incidentally presented items for the different levels of processing conditions in the final free-recall test for each of the four experiments. There was a significant level-of-processing effect on recall of the incidentally presented words in each experiment—Experiment 1, $F(1, 231) = 262.06$, $MSE = 6.726$; Experiment 2, $F(1, 97) = 248.58$, $MSE = 6.229$; Experiment 3, $F(1, 107) = 117.57$, $MSE = 8.376$; Experiment 4, $F(1, 83) = 57.68$, $MSE = 10.597$—with more incidentally presented items recalled in the deep processing condition (Table 2).

False Recall of Critical Intrusions

A 2 × 3 × 2 (Incidental Presentation × Instruction × Level of Processing) mixed ANOVA was computed using the proportion of lists in which critical intrusions occurred as the dependent variable. Instruction was a between-subjects variable in Experiment 1. All of the *F*s for main effects and interactions are reported in Table 3. Most important to Experiment 1, there was a significant three-way interaction mediating all other significant effects. To break down the three-way interaction, we computed simple interaction analyses. Two 2 × 3 (Incidental Presentation × Instruction) ANOVAs were calculated, one for each level of processing condition.

For the deep processing condition, the effects of instruction, $F(2, 111) = 7.27$, $MSE = 0.095$, incidental presentation, $F(1, 111) = 31.93$, $MSE = 0.054$, and the Incidental Presentation × Instruction interaction, $F(2, 111) = 13.75$, $MSE = 0.054$, were all significant (for the incidentally presented condition, $M = 0.368$, and for the nonpresented condition, $M = 0.241$). Simple main effect analyses of instruction at each level of incidental presentation revealed a significant effect of instruction for the incidentally presented (i.e., primed) items, $F(2, 111) = 16.28$, but not for the nonpresented (unprimed) items, $F < 1$. Tukey tests indicate that critical items were recalled significantly more often in the include instruction condition ($M = 0.449$) than in the standard ($M = 0.311$) or exclude ($M = 0.238$) conditions. Furthermore, the critical items were recalled more often in the standard condition than in the exclude condition (see Table 4). These findings show that the baseline level of critical intrusions (unprimed intrusions, or semantic confusion errors) was not significantly affected by instruction. However, episodically caused intrusions were affected by instruction. The difference between the level of semantic confusion errors (intrusions when items were not incidentally presented) and the intrusion level when items were incidentally presented was significant for the include, $F(1, 21) = 33.70$, $MSE = 0.058$, and standard, $F(1, 40) = 7.81$, $MSE = 0.056$, instruction conditions but not for the exclude, $F < 1$, instruction condition (Table 4).

For the shallow processing condition, the main effect of incidental presentation, $F(1, 116) = 14.88$, $MSE = 0.050$, was significant (for incidentally presented, $M = 0.349$, for nonpresented, $M = 0.242$). The effects of instruction, $F(2, 116) = 1.96$, $MSE = 0.121$, and the incidental Presentation × Instruction interaction, $F(2, 116) = 1.16$, $MSE = 0.050$, were both nonsignificant ($M = .352$, .323, and .245 for the include, exclude, and standard instruction conditions, respectively; also, see Table 4).

Discussion

Carefully worded instructions that focused participants on distinctions between incidentally presented items and categorized list words were effective, in the deep processing conditions, for getting participants to bring source memory to bear during the category-cued-recall tests. When told to include category-relevant words from the incidental task, participants increased the frequency of recalling critical items during category-cued recall, and when instructed to exclude incidentally presented items from recall, participants decreased recall of critical items. The effectiveness of the instructions in terms of improving source monitoring depended on the input level of processing of incidentally presented items. Given that they had processed more deeply, which produced better source memory for incidentally presented words, participants were able to intentionally include and exclude incidentally presented category-relevant items on the cued-recall tests. Critical items could not be reliably included or excluded on the category-cued-recall tests if incidentally presented words had been shallowly processed at input.

It is important to note that instructions had no effect on semantic confusion errors; that is, instructing participants to include or exclude incidentally presented items did not reduce critical intrusions in the unprimed condition, with the baseline rate of semantic confusion errors of 24%. This finding extends those of McDermott

Table 3
Summary of Effects of Experiment 1

Source	*df*	*MSE*	*F*
Incidental presentation	1, 227	0.052	45.64*
Instruction	2, 227	0.108	4.85*
LOP	1, 227	0.108	<1
LOP × Instruction	2, 227	0.108	3.47*
Incidental Presentation × Instruction	2, 227	0.052	11.50*
Incidental Presentation × LOP	1, 227	0.052	2.04
Incidental Presentation × Instruction × LOP	2, 227	0.052	3.59*

Note. LOP = level of processing.
* $p < .05$.

Table 4
Means (and Standard Errors) for Primed Critical Intrusions, Unprimed Critical Intrusions, and Episodic Confusion Effects for Experiments 1–3

	Instruction					
	Include		Standard		Exclude	
Experiment	Shallow	Deep	Shallow	Deep	Shallow	Deep
1						
Primed critical intrusions	.443 (.061)	.659 (.069)	.304 (.042)	.384 (.051)	.353 (.046)	.230 (.036)
Unprimed critical intrusions	.261 (.053)	.239 (.048)	.186 (.036)	.238 (.040)	.293 (.047)	.245 (.035)
Episodic confusion effect	.182	.420	.118	.146	.060	−.015
2						
Primed critical intrusions			.360 (.044)	.418 (.043)		
Unprimed critical intrusions			.255 (.041)	.204 (.032)		
Episodic confusion effect			.105	.214		
3						
Primed critical intrusions			.375 (.036)	.415 (.040)		
Unprimed critical intrusions			.265 (.039)	.275 (.036)		
Episodic confusion effect			.110	.140		

Note. Means are the proportions of lists for which critical items were recalled. The episodic confusion effect is the proportion of cases in which primed critical items were recalled minus the proportion of cases in which unprimed critical items were recalled. Shallow and deep refer to level of processing. Only Experiment 1 had include and exclude conditions.

and Roediger (1998), who also found that instructing participants to avoid intrusions did not effectively reduce unprimed false recall.

Significantly more critical intrusions occurred in the include condition than in the standard instruction condition. This finding indicates that the standard instructions were not interpreted by participants the same way as were the include instructions. Of importance, participants in the standard instruction condition did not appear to have intentionally included incidentally presented items on the category-cued-recall tests.

Semantic confusion errors (i.e., intrusions of items that were not primed) were unaffected by the experimental manipulations. The exclude and include instructions affected only episodic confusion errors; as in McDermott and Roediger's (1998) study, semantic confusion errors were not influenced by instructions at recall. Semantic confusion errors were also unaffected by the level of processing of incidentally presented items. As expected, deep processing of incidentally presented words resulted in better free recall of those items in every experiment.

Experiment 2

The results of Experiment 1 show that, given deep processing of incidental words, participants who were instructed to exclude incidentally primed words during recall of the categorized lists could avoid episodic confusion errors (but not semantic confusion errors). If they had shallowly processed the critical items, participants could not have avoided episodic or semantic confusion errors. Do people make such errors if they are not instructed to exclude episodically inappropriate responses? The results of the standard instruction condition in Experiment 1 indicate that participants did commit such errors. Can those errors be identified after the fact? That is, once episodic confusion errors in recall have occurred, is it possible for participants to then identify such errors among their own responses? This question was investigated in Experiments 2 and 3.

The results of the standard instructions condition of Experiment 1, as well as those of a pilot study, indicate that when recall is not accompanied by precisely worded source-monitoring instructions, input level of processing manipulations do not affect the false recall rate of incidentally presented critical items. Deeper input processing might have been expected to enhance retrieval of critical words, thereby increasing false recall. On the other hand, deeper processing should also provide better source memory for critical items, allowing them to be "edited out" from recalled items, because implicitly retrieved responses should not be overtly recalled unless they are also recognized as being members of the appropriate memory set. The improved source memory for deeply processed items should have facilitated this editing process, thereby reducing episodic confusion errors in the deep processing condition. Thus, the lack of a levels of processing effect on false recall might mean either that the two factors (more retrieval but better source memory for critical items) effectively cancelled out each other's effects, or that potential source memory was not spontaneously used during category-cued recall.

It is possible that participants did not bring source memory to bear during the recall test, but that they could do so if they were explicitly asked about the source of their own responses. This reasoning suggests that people might be capable of recognizing, after the fact, when their false memories actually originated from a source other than that of the target memories.

In Experiment 2 of the present study, the effect of level of input processing of critical items on episodic confusion errors was examined. In addition, we examined the ability of participants to recognize post hoc which of their own falsely recalled critical items were actually presented on the incidental task. We hypothesized that deeper input processing of critical items would endow those memory traces with greater source memory potential, allowing for superior source identification of falsely recalled critical items.

Results

False Recall of Critical Intrusions

We computed a 2 × 2 (Incidental Presentation × Level of Processing) mixed ANOVA comparing proportions of lists with critical intrusions as a function of level of processing of the incidental list. There was a significant main effect of incidental presentation on false recall, $F(1, 97) = 21.73$, $MSE = 0.058$; false recall was higher when the critical items had been incidentally presented (Table 4). Neither the effect of level of processing, $F(1, 97) = 0.01$, $MSE = 0.105$, nor the Incidental Presentation × Level of Processing interaction, $F(1, 97) = 2.55$, $MSE = 0.058$, was significant.

The level of episodic confusion errors (Table 4) was calculated as the difference between the level of critical item recall when the items were not incidentally presented (i.e., semantic confusion errors) and the level of critical item recall when critical items were incidentally presented. The level of episodic intrusion errors shown in Table 4 indicates the proportion of critical items falsely recalled because they were incidentally presented.

Source Recognition of Falsely Recalled Incidentally Presented Items

A 2 × 2 (Incidental Presentation × Level of Processing) ANOVA, used as the dependent measure, was computed using the proportion of critical intrusions identified as incidentally presented. There was a significant effect of incidental presentation, $F(1, 41) = 61.29$, $MSE = 0.073$; critical intrusions were more likely to be identified as incidentally presented items when they were incidentally presented (.49) rather than nonpresented (.03). There was also a significant effect of level of processing, $F(1, 41) = 16.99$, $MSE = 0.071$, with better source recognition of falsely recalled incidentally presented items in the deep processing condition (Table 5). The interaction between incidental presentation and level of processing was also significant, $F(1, 41) = 18.34$, $MSE = 0.073$. Simple main effects analyses computed for both shallow and deep processing levels showed that the simple main effect of incidental presentation was significant for both shallow, $F(1, 20) = 5.68$, $MSE = 0.079$, and deep processing, $F(1, 21) = 81.45$, $MSE = 0.067$.

Table 5
Means (and Standard Errors) for Primed and Unprimed Intrusion Source Recognition in Experiments 2 and 3

Experiment	Level of processing	
	Shallow	Deep
2		
Primed intrusion recognition	.242 (.081)	.727 (.077)
Unprimed intrusion recognition	.036 (.026)	.023 (.023)
3		
Primed intrusion recognition	.363 (.061)	.628 (.061)
Remember	.250 (.056)	.507 (.062)
Know	.113 (.039)	.119 (.043)
Unprimed intrusion recognition	.043 (.034)	.014 (.014)

Note. Means are the proportions of critical intrusions recognized as having been on the incidental task.

Discussion

A significant level of episodic confusion errors was found in Experiment 2; more critical words were falsely recalled when those items had been presented on an incidental task, as compared with the incidentally nonpresented condition. As in Experiment 1, in spite of a robust effect of level of processing on final recall of the incidental list, there was not a significant effect of level of processing on the amount of episodic confusion errors in recall. The results of the source-recognition test, however, clearly indicate that deeper processing of the incidental list resulted in better source monitoring for falsely recalled items that were incidentally presented than for critical items that were not presented in the experiment. Although deeper processing yielded better source memory for critical items, as evidenced by the results of the source-recognition test, such source information was apparently not spontaneously used by participants to weed out inappropriate items from the responses they had already falsely recalled. That is, deep processing of the incidental event did not prevent that event from causing false recall. In fact, there was a trend in the direction of more false recall when incidental items were deeply processed, a trend that is opposite to what one would expect if source information had been spontaneously used to carefully monitor recalled words.

Experiment 3

If the task of recognizing the source of one's own episodic confusion errors in the present experiments depends on recollection of the source of incidental events, then factors that improve source memory should affect "remember" judgments rather than "know" judgments when participants try to recognize falsely recalled items from the incidental list among their responses. A "remember" judgment on a recognition memory test refers to the case in which the participant explicitly remembers details about the source of an event, whereas a "know" judgment refers to the case in which the participant has a sense that a memory is familiar but does not explicitly recollect the details of the original event (e.g., Gardiner, 1988). Deeper input processing, which improves recognition memory performance, has been shown to increase the number of "remember" responses without influencing the number of "know" responses on a recognition test (e.g., Gardiner, 1988). To test these predictions, we asked participants in Experiment 3 to indicate whether responses were "remember" or "know" judgments on the test in which they tried to recognize the source of their own episodic confusion errors.

Results

False Recall of Critical Intrusions

A 2 × 2 (Incidental Presentation × Level of Processing) mixed design ANOVA was computed using the proportion of critical items that were falsely recalled as the dependent variable. There was a significant effect of incidental presentation on false recall, $F(1, 107) = 16.83$, $MSE = 0.055$; the intrusion rate was higher when the critical items had been incidentally presented than when they were not (Table 4). Neither the effect of level of processing, $F(1, 107) = 0.39$, $MSE = 0.104$, nor the Incidental Presentation ×

Level of Processing interaction, $F(1, 107) = 0.92$, $MSE = 0.055$, was significant.

Source Recognition of Falsely Recalled Incidentally Presented Items

Participants in the deep processing condition were more likely than those in the shallow processing condition to correctly identify their falsely recalled items that had been presented on the incidental task, $t(88) = 2.94$, $SE = 0.06$. Breaking down the data by "remember" and "know" responses (i.e., the number of critical intrusions that participants remembered or knew were on the incidental task divided by the total number of critical intrusions) showed (Table 5) that significantly more "remember" responses were given in the deep processing condition than in the shallow processing condition, $t(88) = 3.07$, $SE = 0.06$. However, no difference between the deep processing condition and the shallow processing condition was found for the "know" responses, $t < 1$.[2]

Discussion

The results of Experiment 3 replicate and extend the results of Experiment 2. In spite of a level of processing effect on recall of the incidental list, there was not a significant effect of level of processing on episodic confusion errors, which replicates the findings of Experiment 2. Although the input level of processing of critical items did not affect false recall of incidentally presented items, it significantly affected later source recognition of participants' own episodic confusion errors, as we had found in Experiment 2. Furthermore, on the source recognition test, remember source recognition judgments were affected by level of processing, whereas know judgments were not. This finding is consistent with the notion that source memory differed as a function of input level of processing and that this source memory, although encoded and usable, was not spontaneously used to cull falsely recalled items from the correct responses on the category-cued-recall tests. For a somewhat different approach to the phenomenological aspects of false recognition, see Brainerd et al. (2001).

The results of Experiments 1–3 are consistent with the idea that episodic confusion errors in recall occur when automatic retrieval of incidentally presented critical items is not successfully opposed by recollection of those items. In Experiments 2 and 3, there was no benefit of deeper processing on preventing episodic confusion errors in recall. An effect should have occurred in Experiments 2 and 3 if automatic retrieval of critical items was opposed by recollection of the actual source of the critical items (i.e., the incidental task), because there was evidence that deeper processing led to better recollection of incidentally presented items. The tests of source recognition of episodic confusion errors and the final free recall tests showed that deeper processing provided a greater potential for recollection of incidentally presented items. In Experiment 1, it was also shown that instructions for the cued-recall test that encouraged the opposition of automatic retrieval with recollection of incidentally presented items were only effective if critical items had been deeply processed. Shallowly processed items were automatically retrieved, as indicated by the significant episodic confusion effects in recall for those items, but could not be reliably recollected (e.g., on the source-recognition test) in Experiment 1.

Experiment 4

In Experiments 2 and 3, it was demonstrated clearly that participants correctly attributed episodic confusion errors to the incidental list, given deep input processing. What the experiments did not indicate, however, is whether such post hoc corrections of inappropriate memory attributions are accompanied by an escape from the illusion that episodic confusion errors are correct responses from the categorized lists. That is, it may be that participants believed that recalled incidentally presented items were presented on both the incidental task and the categorized lists. This possibility was tested in Experiment 4, in which participants were asked to indicate during the cued-recall tests whether each recalled word had been presented on the incidental list, on the categorized list, or on both lists. We predicted that more accurate identification of episodic confusion errors would be accompanied by a reduction of the belief that incidentally presented items had actually been presented on the categorized lists. We also predicted that deep input processing would better enable participants to identify which critical items had been presented on the incidental list, as we found in Experiments 2 and 3.

Results

Critical Intrusions Attributed to Incidental Task

For each critical intrusion (and for all other recalled words) participants were required to indicate whether they thought the word had been on the incidental list, the studied list, or both lists. If participants realized that a critical item came from the incidental list, they could have attributed the source to either the incidental list or to both the incidental list and the studied list. To see whether level of processing affected the ability to identify the incidental task as at least one source of the critical item, we computed a one-way ANOVA using attribution to the incidental task as the dependent variable. We obtained these attribution scores by adding the proportion of critical intrusions that participants attributed to the incidental list to the proportion of critical intrusions attributed to both lists. Computed in this way, attribution to the incidental task was .56 for the shallow processing condition and .87 for the deep processing condition, a significant effect, $F(1, 75) = 14.31$, $MSE = 0.124$. In the deep processing conditions, relative to shallow processing, participants were more likely to recognize that their critical items came from at least the incidental list (Table 6).

Critical Intrusions Attributed to Study Lists

If participants erroneously believed that a critical item had been on a studied list, they could have attributed its source to either the studied list or to both lists. Even though a "both" response cor-

[2] To determine the levels-of-processing effect on remember and know responses, we analyzed the data in a way suggested by Rajaram (1993): We calculated a ratio of remember responses to the level of overall recognized responses. We then performed paired t tests, comparing the remember-to-recognize ratio in the deep condition with that in the shallow condition. However, following Parkin and Russo (1993), we also conducted an ANOVA in which remember and know responses were combined in a single analysis. The results from the ANOVA agree with those in which remember and know responses are treated separately.

Table 6
Means (and Standard Errors) for Primed Critical Intrusions, Unprimed Critical Intrusions, and Episodic Confusion Effects as a Function of Level of Processing Summed for All Source Assignments for Experiment 4

	Level of processing	
Measure	Shallow	Deep
Primed critical intrusions	.514 (.052)	.740 (.042)
Unprimed critical intrusions	.297 (.048)	.271 (.040)
Episodic confusion effect	.217	.469

Note. Means are the proportions of lists for which critical items were recalled. Standard errors could not be computed for the episodic confusion effect.

rectly identifies the item as a member of the incidental list, the "both" response nonetheless constitutes a memory illusion, because items actually appeared on only incidental lists. To see whether level of processing affected this memory illusion, we computed a one-way ANOVA using false identification as the dependent variable. False identification scores were obtained by adding the proportion of critical intrusions that participants attributed to the studied list to the proportion of critical intrusions attributed to both lists. For the incidentally presented intrusions, the false identification scores were .69 and .56 for the shallow and deep processing conditions, respectively. The effect of level of processing on the memory illusion was not significant, $F(1, 75) = 1.87$, $MSE = 0.166$.

The source-recognition task resulted in many cases in which errors were self-remedied. Primed critical items were recognized as attributable only to the incidental task in 32% of the cases in the deep processing condition and in 44% of the shallow processing condition cases. Table 7 also indicates that when critical items were not incidentally presented, they were almost always (over 90% of the time) attributed exclusively to the study lists once they had been falsely recalled.

Discussion

As in Experiment 1, when participants monitored and judged (i.e., during category-cued recall) the source of each recalled category member online, they were often able to identify critical items that had been presented on the incidental task. Again, this online identification of the true source of episodic confusion errors was significantly better if incidentally presented items were deeply rather than shallowly processed. As in Experiments 1–3, deeper processing appears to have endowed incidentally presented items with more source memory, allowing those items to be properly sourced at recall. The combined results of Experiments 1 and 4 indicate that the correct source of episodically inappropriate responses can sometimes be determined online, as the participant is recalling the list words.

The memory illusion examined in Experiment 4, in which participants believed that critical items had been presented on the studied categorized lists, was powerful for semantic confusion errors (i.e., when critical items were not presented). Participants suffered the illusion approximately 90% of the time when critical items had not been presented. When critical items were incidentally presented, the illusion was considerably weaker (participants suffered the illusion about 60% of the time), but it persisted in spite of source-monitoring instructions or deep processing. Thus, although episodic confusion errors can be remedied to some degree when participants use source memory, such errors nonetheless occur at a high rate.

If participants can correctly determine the source of deeply processed words from the incidental task, why are they nonetheless susceptible to the illusion that the words also occurred on the to-be-remembered lists? One possibility has to do with the diagnosticity of source information or, more specifically, the nondiagnosticity of a lack of source information. Although some words that are correctly recalled from the to-be-remembered lists have appropriate source information, some words are correctly recalled that lack source information identifying them as members of the target list. Given that some words that are correctly recalled lack source information, it is clear that the absence of source information does not diagnose an item as not belonging to a list. For example, a participant recalling the list of birds might correctly recall *sparrow* and falsely recall the incidentally presented word *robin.* If neither memory contained contextual source information, then the source information deficit would not tell the rememberer which was a member of the studied list and which was not. Although the presence of source information identifies an item as a list member, the absence of source information does not identify an item as a nonmember of the study list.

General Discussion

The results of four experiments indicate that false recall can be caused by episodic confusion and that such falsely recalled re-

Table 7
Means (and Standard Errors) for Primed and Unprimed Critical Intrusions by Source Assignment as a Function of Level of Processing for Experiment 4

	Attributed source					
	Incidental task		Studied list		Both lists	
Intrusion	Shallow	Deep	Shallow	Deep	Shallow	Deep
Primed critical	.309 (.074)	.438 (.060)	.438 (.082)	.129 (.038)	.253 (.073)	.433 (.058)
Unprimed critical	.069 (.045)	.034 (.024)	.917 (.046)	.914 (.050)	.014 (.014)	.052 (.038)

Note. Means are the proportions of critical intrusions attributed to each source. Shallow and deep refer to level of processing.

sponses sometimes have the potential to be avoided or even corrected after their commission. In contrast, semantic confusion errors were not self-correctable with source memory, either in the present study or in a study by McDermott and Roediger (1998; but see our Footnote 1). Source memory, improved by deeper input processing in the present experiments, was used by participants to identify their own episodic confusion errors in recall. Episodic confusion errors took the form of falsely recalled responses that are semantically relevant but episodically inappropriate for a to-be-remembered list. The experiments also showed that even when memory accuracy was emphasized, people often did not use available source memory during recall to edit out episodically inappropriate memories from the appropriate ones. Failure to use available source memory during cued recall was shown in Experiments 2 and 3, in which participants often recognized the correct source of their own episodic confusion errors. That is, available source memory was not used to edit out inappropriate responses during category-cued recall, even though accurate source memory was later demonstrated by participants' correct source recognition.

False recall studies have heretofore examined effects of manipulations of level of processing of study lists on false recall, sometimes finding more false recall with deeper processing (Rhodes & Anastasi, 2000) and sometimes showing no effects (Read, 1996). In contrast, in our experiments, we manipulated the level of processing of incidentally primed items rather than target list words. We found no significant effect of level of processing on false recall. Even though the critical intrusion means were slightly greater in the deep rather than in shallow processing conditions in Experiments 2 and 3 and in the standard instruction condition of Experiment 1, the effect sizes (f = .034, .088, and .065 for Experiments 1, 2, and 3, respectively) were all smaller than a small effect size of f = .10 (Cohen, 1988). The lack of levels of processing effects on episodic confusion errors indicates that the extra source memory conferred by deeper input processing was not spontaneously used to edit out episodically inappropriate responses during cued recall. Untapped source memory was found even when incidentally presented words were shallowly processed; about a third of the falsely recalled incidentally presented items that were shallowly processed could be recognized as having been seen on the incidental list. The ability to recognize the correct source of one's own falsely recalled responses was increased substantially by deeper processing of the incidentally presented items words, as shown by Experiments 2 and 3.

The observed failures to spontaneously use available source memory parallel findings by Marsh et al. (1997), who found that unconscious plagiarism of ideas (i.e., cryptomnesia) was reduced only when experimental instructions focused participants' attention on accurate source monitoring. Participants in Marsh et al.'s study who were asked to generate new ideas not previously encountered in the experiment nonetheless plagiarized from ideas generated earlier in the study. When asked to identify the sources of generated ideas, however, participants could often do so, even though they had not used available source memory on the *generate new* task. Other examples of untapped source memory are findings that misinformation effects are mitigated when participants are given a source-monitoring test rather than a simple yes–no recognition test over material (Dodson & Johnson, 1993; Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989). Likewise, Multhaup (1995) found that the false fame effect in both younger and older adults was eliminated when participants were instructed to focus on all possible sources of information when making fame judgments. These examples support Johnson et al.'s (1993) notion that source-monitoring instructions at the time of the test should induce participants to adopt a stricter decision criterion (i.e., attend to the sources of the memories) rather than making recognition judgments with an undifferentiated criterion based on familiarity. The use of familiarity alone to make recognition judgments would be more likely to occur in situations in which specific instructions to monitor the sources of memories are absent.

The episodic confusion errors observed in our experiments may have been caused by automatic retrieval of episodically inappropriate items, unopposed by recollection of the appropriate episodic contexts of the intruded items. This interpretation is based on findings that show that (a) shallow processing of critical items, which limited the ability to recollect those items, produced reliable episodic confusion errors that could not be reliably excluded from recall and (b) deep processing of critical items, which produced the same levels of episodic confusion errors as did shallow processing, nonetheless allowed participants to recognize the source of those items or to exclude those items during cued recall. Thus, when critical items were shallowly processed, automatic retrieval of those items could not be opposed by recollection, and when they were deeply processed, automatic retrieval was not opposed by recollection unless instructions focused on accurate source monitoring.

Although Brainerd et al. (2001) showed that "phantom recollection" (i.e., the false feeling of having recollected a critical nonpresented word) contributed more to false recognition than did familiarity, an alternative account of the present false recall findings is theoretically possible that is based on the familiarity of critical items rather than automatic retrieval of them. According to this explanation, at cued recall, common category members might be implicitly generated from semantic category knowledge and then subjected to a recognition decision. In the absence of specific source-monitoring instructions (which would encourage recollection of source), such recognition decisions might be based primarily on familiarity. The episodic confusion errors observed in the experiments could be due to increased familiarity caused by priming, whether incidentally primed items were shallowly or deeply processed. Furthermore, the effects of instructions in Experiment 1 could be explained in terms of decision criterion shifts; that is, a stricter criterion in the exclude condition reduces extralist intrusions, whereas a more lenient criterion in the include condition increased recall of extralist items. Such an explanation, however, would predict that the stricter criterion in the exclude condition should have also reduced semantic confusion errors, a prediction at odds with the results (see Table 4). This familiarity-based explanation, like the automatic retrieval explanation, addresses the initial recall of intrusions, but it does not explain the effects of the source-recognition tests. The ability participants have to use source memory when directed to do so resembles the ability to use self-generated contextual cues to aid retrieval of associated event memories (e.g., Smith, 1979, 1984). In the case of self-generated context cues, it has been shown that although participants may not spontaneously use nonambient context cues to aid recall, they can effectively do so if, at the time of a recall test, they are instructed to mentally reinstate their learning context. The general pattern indicates that potentially useful source memory is often not spontaneously used to cue and monitor associated event memories.

Although the present results offer some promise that people sometimes may be able to identify their own episodic confusion errors, they do not address false recall that is due to semantic factors such as strength of association (e.g., Deese, 1959; Read, 1996; Roediger & McDermott, 1995) or category membership (Smith et al., 2000). In our study, the baseline condition for false recall of critical nonpresented items was substantial—roughly 25% when critical items were not presented on the incidental task. When recalled critical items were not incidentally presented (semantic confusion errors), they were usually not identified as intrusions. Episodic source information did not appear to reduce such semantically related intrusions.

Participants in false memory experiments are routinely admonished to remember only words from the target list. Likewise, in our experiments participants were given such directions, which were inserted twice in the instructions of all standard conditions. Nonetheless, the results indicate substantial improvements in source monitoring when more specific source-monitoring instructions were given. It cannot be concluded, however, that false recall within the present experimental paradigm can be remedied by creating a greater focus on source monitoring. One limiting factor is the degree of source-memory material encoded during input of the material in question. Memory traces that contain too little episodic source information (e.g., those resulting from shallow processing) are unlikely to be appropriately sourced as a function of instructions, because they do not have the memory potential for adequate memory sourcing. A second factor limiting the conclusions of the present study is the specificity required in the source-monitoring instructions. Participants were not simply asked whether their recalled responses were from any inappropriate source; they were asked specifically about intrusive memories originating from the incidental task. The specificity of these instructions may have allowed participants to retrieve material from memory with cues that were very specific to the episodic source of the incidental task.

It is also important to note a caveat to the conclusion that improved source monitoring can help remedy episodic confusion errors. The results of Experiment 4 show clearly that the memory illusion (i.e., the illusion that an incidentally presented critical item was actually on a studied categorized list) was reduced but not eliminated by focusing participants on accurate source monitoring. Participants believed that more than half of the responses involved in episodic confusion errors in Experiment 4 were presented on the studied categorized lists. Even when participants accurately identified their intrusions as having been on the incidental list, they also believed quite often that the items had also appeared on the study lists. Although instructions can improve the accuracy of source monitoring, our instructions alone did not completely alleviate the faulty source attributions associated with false recall. Furthermore, Hicks and Marsh (2001) found that, as compared with a standard old–new recognition test, false memories were greater on a test that required source monitoring. Thus, monitoring one's source information is by no means a panacea for eliminating false memories.

Results recently reported by Dodhia and Metcalfe (1999) are an interesting contrast to our experiments. The participants in those experiments studied associatively constructed lists (e.g., Deese, 1959) and were given recognition tests over the studied lists. Each inducing list was paired with a second list that sometimes contained the critical linking word. In two experiments, Dodhia and Metcalfe found that false alarms to the critical items were reduced when the critical items were presented in the second list, what they referred to as "inhibition" of the false memory effect. Dodhia and Metcalfe proposed that their participants, aware of the nature of the linking words, used exclusion logic to avoid false alarms in recognition. That is, when participants remembered that the critical items occurred on the alternative list, they reported that the item had not been on the inducing list. At face value, Dodhia and Metcalfe's results appear to contradict our findings, showing not episodic confusion but rather episodic precision. Although the differences between our experiments and those of Dodhia and Metcalfe (1999) include differences in materials (associative vs. categorized lists) and tests (recognition vs. recall), the critical difference appears to be how well camouflaged the critical items were in the two studies. In Dodhia and Metcalfe's study, the inducing and alternative lists were presented as a pair of lists to be differentiated on the source-recognition test. Therefore, the critical items were probably quite salient and meaningful when they appeared on the alternative list immediately following presentation of the inducing list. In our study, however, the incidental presentation of critical items was less salient; critical words appeared simply as items seen earlier on an unrelated incidental task that was not linked to the study lists either temporally or through instructions. The effects of systematic manipulations of the salience of critical items on accurate source identification remain to be tested. It is interesting to note that the salience of the episodically inappropriate items in Dodhia and Metcalfe's experiments did not eradicate the memory illusion; even though participants responded on a recognition test that critical items were not on inducing lists, they often reported during debriefing that it had seemed that the items had actually been on both lists. Thus, in agreement with our results, accurate source memory for the episodically inappropriate events does not eliminate the illusion that items actually had two different sources, the alternative or incidental list and the to-be-remembered list.

Researchers have given much attention to false memories caused primarily by semantic confusion, such as responses that are associatively related to studied items. Errors made when episodic confusion compounds semantic confusion effects comprise an important class of false memories and should be investigated directly.

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629–654.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton and E. Martin (Eds.), *Coding processed in human memory* (pp. 85–124). Washington, DC: Winston.

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

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13,* 207–230.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

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

Dodhia, R. M., & Metcalfe, J. (1999). False memories and source monitoring. *Cognitive Neuropsychology, 16,* 489–508.

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.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review, 4,* 271–276.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

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

Hicks, J. L., & Marsh, R. L. (2001). False recognition occurs more frequently during source identification than during old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 375–383.

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

Jacoby, L. L. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 3–26.

Jacoby, L. L., Woloshyn, V., & Kelly, C. (1987). 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.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17,* 349–358.

Loftus, E. F. (1991). Made in memory: Distortions in recollection after misleading information. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 27, pp. 187–215). San Diego, CA: Academic Press.

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 886–897.

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

McDermott, K. B., & Roediger, H. L., III. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39,* 508–520.

Mulligan, N. W., & Hirshman, E. (1997). Measuring the bases of recognition memory: An investigation of the process-dissociation framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 280–304.

Multhaup, K. S. (1995). Aging, source, and decision criteria: When false fame errors do and do not occur. *Psychology and Aging, 10,* 492–497.

Parkin, A. J., & Russo, R. (1993). On the origin of functional differences in recollective experience. *Memory, 1,* 231–237.

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

Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review, 3,* 105–111.

Rhodes, M. G., & Anastasi, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin & Review, 7,* 158–162.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460–471.

Smith, S. M. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory & Cognition, 12,* 477–482.

Smith, S. M., Ward, T. B., Tindell, D. R., Sifonis, C. M., & Wilkenfeld, M. J. (2000). Category structure and created memories. *Memory & Cognition, 28,* 386–395.

Wells, G. L., & Loftus, E. F. (Eds.). (1984). *Eyewitness testimony: Psychological perspectives.* New York: Cambridge University Press.

Winograd, E. (1968). List differentiation as a function of frequency and retrieval interval. *Journal of Experimental Psychology Monographs, 76* (February, Pt. 2), 1–18.

Zaragoza, M. S., & Koshmider, J. W., III. (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 246–255.

Received May 13, 1998
Revision received August 18, 2000
Accepted August 18, 2000 ■