

Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared

MARA MATHER, LINDA A. HENKEL, and MARCIA K. JOHNSON
Princeton University, Princeton, New Jersey

Subjects hearing a list of associates to a nonpresented lure word later often claim to have heard the lure (Deese, 1959; Roediger & McDermott, 1995). To examine the characteristics of such false memories, subjects completed a memory characteristics questionnaire (MCQ; Johnson, Foley, Suengas, & Raye, 1988) or made remember/know (RK; Gardiner & Java, 1993) judgments for previously heard theme associates and nonpresented lures. MCQ ratings indicated that false memories for lures had less auditory detail and less remembered feelings and reactions than memories for presented words. In addition, rates of false recognition for lures were significantly lower than rates of correct recognition when items from various themes were intermixed instead of blocked at acquisition and subjects made MCQ ratings instead of RK judgments. This demonstrates that false memories can be affected both by how they are acquired and by how extensively they are examined at retrieval.

Memory is rarely exact, but it is usually not completely wrong. When remembering a friend's comment or a passage from a book, we are likely to remember the general ideas expressed but not the exact words used (e.g., Bransford & Franks, 1971; Sachs, 1967). To fill in the gaps, we often assume, infer, or imagine what happened. However, these internally generated events, ideas, or beliefs sometimes go beyond filling in minor gaps, creating memories for things that we never experienced (e.g., Bartlett, 1932; Bransford & Johnson, 1973). Such false memories include remembering experiencing events that we only imagined (e.g., Johnson & Raye, 1981), had suggested to us (e.g., Loftus, 1979), or inferred on the basis of our prior knowledge or schemas (e.g., for reviews, see Alba & Hasher, 1983; Ceci & Bruck, 1993; Johnson, Hashtroudi, & Lindsay, 1993).

In certain situations, such as when giving or hearing eyewitness testimony, it is critical to distinguish false from true memories. Sometimes false memories do differ in subtle ways that indicate that they seem less vivid and inspire less confidence for the rememberer. For example, when people incorrectly identify a suspect in a lineup, they take longer, are less confident (Sporer, 1992; Sporer, Penrod, Read, & Cutler, 1995), and report more explicit cognitive effort (Stern & Dunning, 1994) than when they

make a correct identification. When people claim to have seen a picture of something they had actually seen only the name of, they indicate less conscious recollection of having seen the item as a picture than when they had actually seen a picture of the item (Lane & Zaragoza, 1995). Also, when recalling nonexistent objects after being exposed to misleading postevent information, people provide longer descriptions with more hedges, more references to cognitive operations, and fewer sensory details than when recalling objects that were actually present (Schooler, Gerhard, & Loftus, 1986).

However, some false memories seem indistinguishable from true memories. For example, in some situations, subjects who are misled (e.g., they see a hammer in a sequence of slides, but afterwards read a narrative in which it is referred to as a wrench) actually express more confidence when judging misled items as old than when judging old items as old (Loftus, Donders, Hoffman, & Schooler, 1989, Experiment 2) and even bet equivalent amounts of money on suggested and actual memories they called old (Weingardt, Toland, & Loftus, 1994). Does such confidence reflect a situation in which the false memory actually has the same characteristics as those of a true memory? More generally, what determines whether true and false memories can be distinguished?

Addressing this problem, Johnson and her colleagues (e.g., Johnson, 1988; Johnson et al., 1993; Johnson & Raye, 1981) proposed a source-monitoring framework (SMF) for characterizing the qualitative characteristics of memories (both true and false) from various sources (e.g., seeing, hearing, imagining, or reading about events) and the factors affecting whether source misattributions will occur. They argue that both accurate and inaccurate remembering result from an interaction between activated information and judgment processes. Findings from a number of

This research was supported by National Institute on Aging Grant AG09253 and a National Science Foundation Graduate Fellowship awarded to the first author. Thanks are extended to Jennifer Glaza and Christine Voegeli for assistance in coding the data, Scott Nolde and Christine Voegeli for recording the stimuli, and Robert Belli, Jonathan Schooler, and the 1995-1996 Princeton Cognition Lab members for helpful comments. Correspondence should be addressed to M. Mather, Department of Psychology, Princeton University, Princeton, NJ 08544-1010 (e-mail: tmather@princeton.edu).

experimental paradigms, for example, source discrimination, misleading suggestion, and false fame, are consistent with the SMF (for a review, see Johnson et al., 1993; for application to eyewitness testimony, see Belli & Loftus, 1994, Lindsay, 1994, Zaragoza & Lane, 1994; for application to recovery of repressed memories, see Lindsay & Read, 1994).

In general, according to the SMF, memories from different sources—for example, imaginal/inferential sources versus perceptual sources—normally differ in their average qualitative characteristics. These typical average differences can be used to judge the source of a memory. The more that particular memories from different sources overlap in their various characteristics, the more difficult it will be to correctly attribute them to their original source. Characteristics that are important for determining source include information about the cognitive operations that created a mental experience and its semantic, perceptual, contextual, and affective detail. Thus, for example, internally generated words that require little cognitive effort to generate are more likely to be later judged as having actually been presented than are those that require more effort, because, typically, perceived events require less reflectively guided operations than do self-generated events (Finke, Johnson, & Shyi, 1988; Johnson, Raye, Foley, & Foley, 1981). Similarly, increasing the semantic (Lindsay, Johnson, & Kwon, 1991), perceptual (Johnson, Foley, & Leach, 1988), or both semantic and perceptual (Henkel & Franklin, in press) similarity between memories from various sources reduces source accuracy.

In addition to factors that should affect the characteristics of mental experiences, the SMF proposes that the type and parameters of judgment processes engaged at test should affect the accuracy of source attributions as well. For example, people sometimes use plausibility or the extent to which information fits with a schema to discriminate between sources, especially when simple characteristics of the memory are ambiguous as to source. If information fits with prior knowledge or beliefs, individuals may examine memories less carefully (Johnson, 1988; Johnson, Foley, Suengas, & Raye, 1988). Making source decisions using a lenient criterion (e.g., of familiarity) is likely to result in more false memories than making source decisions using more stringent criteria (e.g., amount of perceptual and contextual detail) (see, e.g., Dodson & Johnson, 1993, 1996; Jacoby, Kelley, Brown, & Jasechko, 1989; Lindsay & Johnson, 1989; Raye, Johnson, & Taylor, 1980; Zaragoza & Koshmider, 1989).

A recently revived list-learning paradigm (Deese, 1959; Read, 1996; Roediger & McDermott, 1995) offers another context in which to investigate the question of the relation between the characteristics of true and false memories using a simple task that results in a high rate of false recognition. In this paradigm, subjects hear several lists of words that are highly associated with a word that is not presented (e.g., *thread, pin, eye, sewing, sharp, point, prick, thimble, haystack, thorn* for the critical lure *needle*). They are likely to erroneously recall or recognize

the lure, often at the same rate as for presented words, and during the debriefing it can be difficult to convince them that they did not hear the lure (see, e.g., Payne, Elie, Blackwell, & Neuschatz, 1996).

In the Deese paradigm, activation of lure words is likely to occur relatively automatically, without specific reflective operations. Hence, failure to correctly identify whether a lure was perceived or was internally generated (i.e., a reality monitoring error) would be expected on these grounds (e.g., Johnson et al., 1981). Furthermore, according to the SMF, we would expect both acquisition and test conditions to affect the rate of false memories in this paradigm. Conditions that increase the chance that a lure will be generated at acquisition should increase later false recognition of the lure (e.g., Durso & Johnson, 1980). Also, conditions that increase the chance that information will be encoded in terms of a common schema, model, or image should increase the chances that schema-relevant false recognitions will occur (e.g., Johnson, Bransford, & Solomon, 1973; Sulin & Dooling, 1974). Consistent with these expectations, a study using the Deese (1959) paradigm showed that the number of lures incorrectly recalled is affected by whether the words were organized in blocks of related items (McDermott, 1996). On immediate recall, participants who heard all the items associated with a lure sequentially (a blocked condition) were more likely to mistakenly recall those lures rather than studied items, but subjects who heard all the associated items intermixed throughout the study session (a random condition) recalled lures and studied items at the same rate. Blocking items very likely increases both the chances that the lure will be activated at encoding and the chances that an overall list schema will be induced by making the theme of each set more apparent.

In short, the high rates of false recognitions in the Deese paradigm indicate that true and false memories must be similar in that context. However, the high error rates do not necessarily mean that true and false memories are alike in every respect. Errors simply reflect that false and true memories are similar on the dimensions that are actively being used to make source judgments; these judgments are made either heuristically or more systematically, drawing on inferential/reasoning processes. For example, because there is likely to be high overlap between semantic/associative features of perceived and associatively activated items, semantic detail should not be a particularly diagnostic cue for source. Source judgments that heavily weight semantic information would then result in a high rate of source misattributions. Nevertheless, true and false memories might differ on other characteristics (e.g., perceptual detail) that might not be so salient to subjects in this situation, but which could be used under some circumstances to distinguish source.

The present study was designed to explore two questions: First, do true and false memories in this situation have the same qualitative characteristics? Qualitative features of memories were assessed using a modified memory characteristics questionnaire (MCQ; Johnson, Foley, et al., 1988; Johnson, Nolde, & De Leonardis, 1996). An

MCQ assesses subjective ratings of specific characteristics of memories (e.g., perceptual or emotional detail). Of interest were which specific characteristics of the false memories contributed to how realistic the memories seemed. In particular, do true and false memories differ more in the amount of associated perceptual detail (the studied words were heard and the lures were not) than in the amount of remembered cognitive operations or associations (the critical lures were likely to have been generated effortlessly and reactivated along with presented items)?

A way phenomenological qualities of memories in this paradigm have been investigated previously is by asking subjects to judge the relative extent to which they consciously recollect the experience of hearing the word, using the remember/know (RK) distinction (Gardiner & Java, 1993; Tulving, 1985). *Remember* judgments are for vivid, consciously remembered memories, and *know* judgments are for confident memories that have no associated specific detail. The proportion of *remember* judgments for critical lures in the Deese (1959) paradigm tends to be quite high. *Remember* judgments for lures are higher than *know* judgments and are sometimes as high as *remember* judgments for studied items (Payne et al., 1996; Roediger & McDermott, 1995; but see Norman & Schacter, 1997; Read, 1996), indicating that the false memories seem almost as vivid as the true memories.

However, one piece of evidence suggesting that false memories of lures are not identical to veridical memories of studied items is that subjects were most likely to recall the lure toward the end of their output (Payne et al., 1996; Roediger & McDermott, 1995). Therefore, it is possible that the memories may differ on some specific dimensions that may not always be reflected in the more general *remember* ratings. That is, both true and false memories may have qualities vivid enough to assign them *remember* ratings, but which qualities are vivid for each type of memory might be different. For example, true and false memories might both have vivid semantic/associative qualities, but only the true memories might have vivid perceptual qualities. We therefore obtained MCQ ratings to more directly assess the relative qualitative characteristics of memories for studied items and lures.

The second question we explored was whether individuals could be induced to evaluate their memories more critically and thus decrease false recognitions in this situation. For example, with a *yes/no recognition test*, Lindsay and Johnson (1989) replicated the standard finding that subjects were likely to falsely recognize information introduced through misleading suggestion subsequent to a target event; in contrast, subjects were less likely to falsely recognize suggested information with a *source-monitoring test*. Evidently, the source-monitoring test induces subjects to evaluate memories more carefully (also see Dodson & Johnson, 1993; Multhaup, 1995; Zaragoza & Koshmider, 1989). In the present study, we compared two test conditions: RK and MCQ. In both conditions, subjects were required to make source judgments (which of two speakers said each word); thus,

evaluation criteria would be expected to be relatively high in both conditions. Nevertheless, the RK instructions simply ask subjects whether they remember any specific detail, whereas the MCQ asks about several types of qualitative information (perceptual, affective, semantic/associative, overall confidence). Thus, even though both groups are engaged in source monitoring, the MCQ might induce a still more stringent evaluation of memories than would the RK judgment task, either by increasing the number of characteristics considered or by increasing the amount of information required to call an item old.

In summary, to investigate these issues regarding the phenomenal similarity and differences between false and veridical memories induced by the Deese (1959) list paradigm, we examined the effects of studying the words in a blocked (the words associated with each lure heard sequentially) or random (the words associated with each critical lure heard interspersed throughout the tape) format, with either one or two speakers saying all the words associated with a critical lure. To assess the resulting phenomenological characteristics of the memories, we used two measures at test. Some subjects completed the MCQ for test items and others made RK judgments about each word they claimed to have heard.¹ The RK test also allows us to make a direct comparison with other related studies. We expected that subjects might assign *remember* judgments as often to lures as to studied items in the blocked conditions (as in Payne et al., 1996), but that *remember* judgments might be less likely in the random conditions. We also expected that subjects who were required to examine particular aspects of their memories (MCQ condition) rather than make a more global judgment (RK condition) might be less likely to report having heard words they had not heard (e.g., Dodson & Johnson, 1993; Lindsay & Johnson, 1989).

METHOD

Subjects

Princeton University undergraduates (58 women and 38 men, with a mean age of 20 years) participated and either received course credit or payment. One additional student participated but is not included in any analyses because of a failure to follow instructions.

Materials

Fifteen 10-word sets were selected from twenty-four 15-word sets (Roediger & McDermott, 1995). Each set was abbreviated by omitting the last five words. All 10 words in each set were highly associated with a lure that was not presented during the study session. The sets were divided into three groups of five sets for counterbalancing purposes. Across subjects, each group (and therefore each word from each list) was assigned to old and new categories equally often. The study lists were constructed from a single pre-recorded master list of these sets of words.

Each subject was presented with the 10 words from 10 of the sets during the study session (yielding a 100-word study session). In the blocked condition, all 10 associates for each critical lure were presented in one list, with the strongest associates occurring first. In the random condition, the 10 items on each list were drawn from the 10 different study sets; thus, each item in a presentation list corresponded to a different lure. In all conditions, subjects heard both a

male speaker and a female speaker say the words. In both the blocked and random conditions, half the subjects heard each speaker saying all the words for half the theme sets (one-speaker condition), and the other half heard five of the words in a theme set said by the male speaker and five said by the female speaker (two-speaker condition). In all conditions except the one-speaker blocked condition, no more than two words in a row were heard spoken by the same speaker. The words were presented at the rate of 2 sec per word. After each list of 10 items, there was a click sound, a 5-sec pause, then the phrase “next list” before the next list.

The recognition test consisted of 4 items from each 10-word presentation list (in each case, the items that had been presented in Positions 1, 4, 6, and 8 in each list) in addition to the 10 lures. To keep the recognition test the same for the random and the blocked conditions, if a word had been in one of the four tested positions on a blocked list (1, 4, 6, or 8), it was in one of those four positions in one of the random lists as well. Because the subject had not heard the items from five of the sets, these items (and their five related lures) served as new items. The total number of test items was 75.

Procedure

A 2 (format: blocked, random) × 2 (speaker per theme: one, two) × 2 (test: RK, MCQ) factorial design was used, with 12 subjects randomly assigned to each condition. The subjects were tested individually or in groups of 2 or 3. They were told that they would hear lists of words on a tape recorder to see if the number of voices reading words affected their memories. They were told that they would hear word lists read by 2 people (“Steven” and “Cathy”) and to try to remember what they heard. They were not told to remember who said each word.

After hearing the tape, the subjects were given written instructions for the memory test. Half of the subjects were given the RK and half were given the MCQ questionnaire. All subjects were asked to judge whether each word had been said by Steven or Cathy, or was a new word that they had not heard on the tape. The subjects in the RK condition made a *remember* or *know* judgment for each word they decided Steven or Cathy had said. The instructions for the RK decisions were modeled after Rajaram’s (1993) instructions.

The subjects were instructed to use a *remember* response when they could consciously recollect details of the actual occurrence of the word and to use a *know* response when they were confident that the item was presented on the tape, but could not remember anything specific about its presentation. We did not collect confidence ratings in the RK condition, because that might lead subjects to confuse the RK judgment task with a confidence-rating task.

After making the Steven, Cathy, or new judgment, subjects in the MCQ condition evaluated their memories using the MCQ for each word they decided Steven or Cathy had said. First, for each response, they rated their confidence (they were asked to “indicate how confident you are that your response is correct”). Then, for the words they classified as old, they rated the amount of auditory detail, amount of information about the speaker’s or their own feelings or reactions, the number of associations to that word,² and the extent to which they recalled thinking about the word during the study session in an effort to remember it. They were instructed that confidence about a word was not necessarily based on the total amount of detail available to them in memory—that at times they might be very confident that a word had been presented but not remember any specific details of its presentation.

RESULTS

Recognition

Table 1 shows the mean proportion (and standard error of the mean) of studied items, critical lures, and new items called old³ (items that subjects said were spoken by either Steven or Cathy) for both the RK and the MCQ conditions. All subsequent comparisons among means were planned, an alpha level of .05 was used for all statistical tests, and confidence intervals of ±2SE around each mean are given after each mean.

As expected, the blocked condition ($M = .78 \pm .04$) produced higher old responses overall to studied and lure items than did the random condition ($M = .67 \pm .06$). This

Table 1
Proportion of Items Judged to Be Old at Test and
Proportion of Correct Source Attributions for the RK and MCQ Conditions

| Study Condition | RK Condition | | | | MCQ Condition | | | |
|---------------------|---------------------|-----------|--------------------------------|-----------|---------------------|-----------|--------------------------------|-----------|
| | Proportion Said Old | | Proportion Correct Attribution | | Proportion Said Old | | Proportion Correct Attribution | |
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Studied | | | | | | | | |
| Blocked one-speaker | .76 | .04 | .84 | .03 | .83 | .02 | .84 | .06 |
| Blocked two-speaker | .75 | .04 | .65 | .03 | .73 | .04 | .65 | .04 |
| Random one-speaker | .69 | .04 | .69 | .04 | .65 | .04 | .72 | .04 |
| Random two-speaker | .75 | .03 | .63 | .04 | .79 | .04 | .59 | .03 |
| Average | .74 | .02 | .71 | .02 | .75 | .02 | .70 | .02 |
| Critical Lure | | | | | | | | |
| Blocked one-speaker | .80 | .05 | .86 | .03 | .85 | .03 | .80 | .06 |
| Blocked two-speaker | .73 | .05 | | | .80 | .05 | | |
| Random one-speaker | .62 | .07 | .66 | .06 | .52 | .11 | .72 | .06 |
| Random two-speaker | .76 | .05 | | | .56 | .08 | | |
| Average | .73 | .03 | .76 | .04 | .68 | .04 | .76 | .04 |
| New | | | | | | | | |
| Blocked one-speaker | .20 | .05 | | | .24 | .06 | | |
| Blocked two-speaker | .15 | .03 | | | .19 | .06 | | |
| Random one-speaker | .33 | .06 | | | .29 | .05 | | |
| Random two-speaker | .33 | .05 | | | .30 | .07 | | |
| Average | .25 | .03 | | | .26 | .03 | | |

difference was reflected in a 2 (format: blocked, random) \times 2 (speaker per theme: one, two) \times 2 (test: RK, MCQ) \times 2 (item type: old, lure) mixed-factor analysis of variance (ANOVA), which revealed a main effect of format [$F(1,88) = 12.09, MS_e = .05$]. Also, there was a significant format \times speaker per theme interaction [$F(1,88) = 5.24, MS_e = .06$], which indicated that subjects in the one-speaker condition were more likely to call studied items and lures old if they had heard the study words in a blocked format than if they had heard them in a random format (mean difference = $.19 \pm .10$), whereas in the two-speaker condition the format did not significantly affect old responses (mean difference = $.04 \pm .09$). In addition, there was a main effect of item type [$F(1,88) = 4.59, MS_e = .02$] and a significant format \times item type interaction [$F(1,88) = 13.05, MS_e = .02$]. These effects were both qualified by a significant format \times test \times item type interaction [$F(1,88) = 6.61, MS_e = .02$]. To explore the nature of this three-way interaction, we analyzed the MCQ and RK conditions separately (see Table 2). As evident in Table 2, the general pattern was similar in the MCQ and RK conditions. There was little difference between the studied items and lures in the proportion of items called old in the blocked conditions; in the random conditions, fewer lures than studied items were called old. However, the interaction between format and item type was independently significant only in the MCQ condition [$F(1,44) = 16.64, MS_e = .02$]. In planned comparisons for both the RK and the MCQ conditions, the MCQ random condition was the only condition in which significantly fewer lures than studied items were called old ($p < .0001$).

In summary, we found no difference in the proportion of items called old for studied items and lures in most conditions, in line with previous studies that had found that critical lures were called old at rates equivalent to or higher than those for studied items (Payne et al., 1996, Experiment 1; Roediger & McDermott, 1995, Experiment 2). However, lures were less likely to be called old in the MCQ random condition. This indicates that factors at both encoding and retrieval can affect the level of false recognition. At encoding, the presentation of items in a blocked format leads to high rates of false memories. These false memories are quite resistant to scrutiny, persisting even when several specific characteristics are examined. However, false memories created in the less structured random condition are more vulnerable to the comprehensive examination encouraged by the MCQ condition.

Table 2
Proportion of Critical Lures and Studied Items Identified as Old in Blocked and Random Conditions for RK and MCQ Conditions

| | RK | | | | MCQ | | | |
|----------------|----------|-----------|----------|-----------|----------|-----------|----------|-----------|
| | Blocked | | Random | | Blocked | | Random | |
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Studied items | .76 | .03 | .72 | .03 | .78 | .03 | .72 | .03 |
| Critical lures | .77 | .03 | .69 | .05 | .82 | .03 | .54 | .07 |

Source Monitoring

The proportions of studied items called old that were attributed to the correct speaker are presented in Table 1 for each condition. As can be seen in the table, source monitoring in our experiment was most accurate when one speaker said all the items related to a theme and all related items were in one block. In addition, source-monitoring rates for studied items in our experiment were quite similar to levels in a previous experiment (Payne et al., 1996, Experiment 3), with .84 of the items correctly attributed in a one-speaker blocked condition in that study (compared with our .84 for both the MCQ and the RK conditions) and .71 in a condition similar to our two-speaker blocked condition (which yielded a source monitoring rate of .65 across MCQ and RK).⁴

Table 1 also shows the proportions of lures called old that were attributed to the "correct" speaker, that is, the speaker who said all the words in the list related to the critical lure. In the two-speaker conditions, half the old items associated with each lure were spoken by each speaker, so there is no logical "correct" source and we do not present source-monitoring data for those lures. Across lures and studied items, source attribution was better in blocked conditions ($M = .84 \pm .04$) than in random conditions ($M = .70 \pm .04$), indicating that the blocked format helped subjects associate words with the appropriate speaker. That the subjects were using the structure of the study lists to help them discriminate who said each word was also indicated by the fact that they were as likely to claim that the speaker who said all the words in a list had said the corresponding critical lure ($M = .76 \pm .04$) as to correctly identify the speaker who said a word they had actually heard ($M = .77 \pm .06$).

It is interesting that in the Payne et al. (1996) study, even though subjects were quite willing to attribute the lures to a speaker (87% were attributed compared with 94% for studied items), they were not much more accurate than chance (in the condition corresponding to our one-speaker blocked condition, they were "correct" only 53% of the time). The different levels of source monitoring may be due to differences in the structure of the two experiments. In the Payne et al. study, format was a within-subjects manipulation, and thus each subject heard only two theme sets for which the words were all spoken by the same speaker. In contrast, in our one-speaker conditions (in which the words for every theme set were spoken by just one speaker), subjects may have been more likely to use the structure of the study lists to help infer who said a word. In addition, the words in the Payne et al. study were presented by speakers on a videotape, which may have encouraged focusing on remembered perceptual detail to discriminate the source. Judgments based on perceptual characteristics would lead to accurate source judgments for old words but not for lures. In contrast, in our study, the structured lists and the lack of perceptual detail may have led associations to be the most salient cue in discriminating who said what items. Unlike relying more on perceptual detail, using associations to discriminate who said what words would be likely to lead to

source-monitoring performance that was as good for lures as for old items (and potentially even better for lures, since the lures were associated with every target word, whereas the target words were not necessarily associated with each other).

In summary, when Speaker A said all the associated items for a particular lure, that lure was more often attributed to Speaker A than to Speaker B. This indicates that false memories for lures are affected by the characteristics of the memories for associated studied items, either through inference processes at test or when associations to the lure come to mind during acquisition.⁵

Phenomenal Qualities of Memories

Remember/know judgments. The results from the RK task are presented in Table 3 by item type and study condition. *Remember* and *know* responses were calculated as proportions of all old responses given each type of response (so that the *remember* and *know* responses add up to 100 except in conditions in which one or two RK judgments were left blank).⁶ There were no significant differences between the proportions of *remember* and *know* responses for studied items. Lures were judged to be known ($M = .69 \pm .10$) more often than remembered ($M = .31 \pm .10$), in contrast to previous studies using this experimental paradigm, in which subjects made equivalent or more *remember* than *know* responses for critical lures (Payne et al., 1996; Roediger & McDermott, 1995). However, in the one-speaker blocked condition (which most closely corresponds with the format of the previous studies), the proportion of *know* ($M = .56 \pm .20$) and *remember* ($M = .42 \pm .20$) judgments to lures were more similar to each other.

Subjects were more likely to give a *remember* response when they said a studied item was old ($M = .48 \pm .06$) than when they said a lure was old ($M = .31 \pm .10$). How-

ever, this difference was less pronounced if we considered just the *remember* ratings for lures ($M = .42 \pm .20$) and old items ($M = .50 \pm .14$) in the one-speaker blocked condition. The fact that rates of *remember* judgments were generally lower for lures than for studied items in this study is not surprising, given that they differ in some other studies as well (Norman & Schacter, 1997; Read, 1996). However, it is interesting that the rates of *remember* judgments do not always differ between lures and old items (Payne et al., 1996; Roediger & McDermott, 1995). It is not clear what may lead to the different pattern of ratings across studies, especially because the different pattern emerges in two experiments using the same stimuli and procedure (Norman & Schacter, 1997; Roediger & McDermott, 1995).

In general, the rates of *remember* responses were lower for studied items than in previous studies. It is possible that the *remember* judgments were generally less frequent in this study than in previous studies using this paradigm, because here subjects made source judgments before each *remember/know* judgment, whereas they just made old/new judgments in the other studies (cf. Lindsay & Johnson, 1989; Multhaup, 1995). Because they were required to identify who said each item they remembered, they might have been less likely to choose a *remember* response for the items for which they could not remember the speaker. Indeed, when we compared the proportions of *remember* responses for studied items that were attributed to the correct speaker with those attributed to the wrong speaker, subjects were more likely to make a *remember* judgment for correctly attributed words ($M = .55 \pm .06$) than for incorrectly attributed words ($M = .37 \pm .08$).

In contrast, a parallel comparison for the lures in the one-speaker conditions revealed no difference in the proportion of *remember* responses for lures that were “correctly” attributed ($M = .37 \pm .16$) and those that were “incorrectly” attributed ($M = .36 \pm .14$), even though subjects were about as “good” at source monitoring for the lures as for the studied items. Thus, subjects were relatively unlikely to use a *remember* judgment for the lures, whether or not they “correctly” attributed it. This suggests that the higher levels of *remember* judgments for correctly attributed words were due to greater associated perceptual detail, not just to greater certainty about the speaker. A similar pattern of results was obtained in a study in which subjects made source monitoring judgments about actions that they had either performed themselves, watched being performed, or imagined being performed (Conway & Dewhurst, 1995). Subjects gave a higher proportion of *remember* responses when they correctly attributed performed or observed actions than when they incorrectly attributed such actions, but showed no difference in the proportion of *remember* responses for correctly and incorrectly attributed imagined actions. Thus, in the context of a source-monitoring task, *remember* responses seem to reflect elevated perceptual information and are not simply associated with correct attributions. Additionally, source monitoring seems to induce stricter criteria

Table 3
Proportion of *Remember* and *Know* Responses
for Items Judged to Be Old

| Study Condition | <i>Remember</i> | | <i>Know</i> | |
|---------------------|-----------------|-----------|-------------|-----------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Studied | | | | |
| Blocked one-speaker | .50 | .07 | .49 | .07 |
| Blocked two-speaker | .41 | .05 | .59 | .05 |
| Random one-speaker | .47 | .05 | .53 | .05 |
| Random two-speaker | .51 | .06 | .49 | .06 |
| Average | .47 | .03 | .53 | .03 |
| Critical Lure | | | | |
| Blocked one-speaker | .42 | .10 | .56 | .10 |
| Blocked two-speaker | .20 | .08 | .79 | .08 |
| Random one-speaker | .29 | .09 | .71 | .09 |
| Random two-speaker | .32 | .08 | .68 | .08 |
| Average | .31 | .05 | .69 | .05 |
| New | | | | |
| Blocked one-speaker | .44 | .17 | .54 | .15 |
| Blocked two-speaker | .13 | .07 | .84 | .07 |
| Random one-speaker | .10 | .07 | .90 | .07 |
| Random two-speaker | .34 | .11 | .66 | .11 |
| Average | .25 | .05 | .74 | .05 |

for *remember* ratings, leading to comparatively fewer *remember* ratings for incorrectly attributed old items and lures than found in other studies using the Deese paradigm.

Memory characteristics questionnaire. With the MCQ, we were able to more finely explore the phenomenological differences between actual and false memories suggested by the lower rates of *remember* judgments for false than for veridical memories. Table 4 shows the MCQ subjects' confidence scores. In addition, the mean MCQ ratings for each condition are shown for the amount of remembered auditory detail, the subject's or the speaker's feelings/reactions, associations, and trying to remember. Confidence was rated on a 3-point scale (1 = *guessing*, 2 = *fairly sure*, and 3 = *very sure*); auditory detail and the remaining items on the MCQ were rated using a 5-point scale, with 1 = *little or none of that feature* and 5 = *feature clear or distinct*.

Studied items were recognized with more confidence ($M = 2.15 \pm .10$) than were lures ($M = 2.02 \pm .12$). The mean difference between ratings for old words and critical lures was significant for auditory detail (mean difference = $.41 \pm .19$) and feelings/reactions (mean difference = $.22 \pm .19$) but not for associations (mean difference = $.08 \pm .19$) and trying to remember (mean difference = $.05 \pm .16$). Both auditory detail and reactions to an event act as contextual detail and can potentially help identify an event as externally generated. That is, memories for perceived events typically include more perceptual detail and more information about thoughts and feelings than do memories for imagined events (Hashtroudi, Johnson, & Chrosniak, 1990; Hyman & Pentland, 1996; Suengas & Johnson, 1988).

As noted above, *remember* judgments were higher for old items that were correctly attributed than those incorrectly attributed, but there was no such difference for the lures. The same pattern emerged for the MCQ judgments. Each characteristic (including confidence) was rated significantly higher for correctly attributed than for incorrectly attributed memories of studied items, suggesting that studied items that were incorrectly attributed to a speaker had less distinctive features in general than did correctly attributed studied items. The mean differences and confidence intervals were $.42 \pm .21$ for auditory detail, $.34 \pm .24$ for feelings/reactions, $.36 \pm .19$ for associations, and $.24 \pm .21$ for trying to remember. This result is consistent with the higher levels of *remember* judgments found for correctly attributed words in the RK condition, and furthermore suggests that elevated contextual information, which is more likely to be present with old items than with lures, may lead to *remember* judgments.

A parallel analysis was conducted for the characteristics of "correctly" and "incorrectly" attributed lure items (in the one-speaker conditions). Lures that were "correctly" attributed were given higher confidence ratings (mean difference = $.53 \pm .39$). In addition, the "correct" attributions were rated as having significantly more associations (mean difference = $.70 \pm .57$). However, none of the other characteristics differed significantly between "correct" and "incorrect" attributions. This indicates that attributing the lures to the "correct" source is associated with a greater sense of confidence that the lure was heard, but not with greater vividness of associated contextual details in general. In addition, the higher level of related associations for the "correctly" attributed lures

Table 4
Average MCQ Ratings

| Study Condition | Confidence | | Auditory Detail | | Feelings/Reactions | | Associations | | Trying to Remember | |
|---------------------|------------|-----------|-----------------|-----------|--------------------|-----------|--------------|-----------|--------------------|-----------|
| | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> | <i>M</i> | <i>SE</i> |
| Studied | | | | | | | | | | |
| Blocked one-speaker | 2.28 | .11 | 2.81 | .27 | 2.64 | .24 | 3.16 | .21 | 2.13 | .26 |
| Blocked two-speaker | 2.04 | .08 | 2.57 | .22 | 2.26 | .15 | 3.31 | .27 | 2.24 | .28 |
| Random one-speaker | 2.17 | .10 | 2.62 | .21 | 2.37 | .23 | 2.59 | .22 | 2.29 | .26 |
| Random two-speaker | 2.12 | .09 | 2.79 | .23 | 2.53 | .23 | 2.37 | .26 | 2.04 | .30 |
| Average | 2.15 | .05 | 2.70 | .11 | 2.45 | .11 | 2.86 | .13 | 2.17 | .14 |
| Critical Lure | | | | | | | | | | |
| Blocked one-speaker | 2.32 | .11 | 2.57 | .31 | 2.72 | .27 | 3.57 | .26 | 2.11 | .27 |
| Blocked two-speaker | 1.89 | .09 | 2.12 | .19 | 2.03 | .17 | 3.09 | .28 | 2.20 | .28 |
| Random one-speaker | 1.99 | .15 | 1.85 | .24 | 1.90 | .31 | 2.29 | .31 | 2.35 | .34 |
| Random two-speaker | 1.90 | .13 | 2.53 | .26 | 2.19 | .30 | 2.25 | .33 | 1.94 | .24 |
| Average | 2.03 | .06 | 2.27 | .13 | 2.21 | .14 | 2.80 | .17 | 2.14 | .14 |
| New | | | | | | | | | | |
| Blocked one-speaker | 1.64 | .11 | 2.08 | .31 | 2.13 | .26 | 2.51 | .37 | 1.83 | .37 |
| Blocked two-speaker | 1.56 | .13 | 2.36 | .33 | 2.54 | .40 | 2.75 | .38 | 2.07 | .29 |
| Random one-speaker | 1.59 | .15 | 1.82 | .23 | 1.66 | .24 | 1.53 | .20 | 1.71 | .30 |
| Random two-speaker | 1.52 | .15 | 2.16 | .19 | 1.91 | .27 | 1.65 | .22 | 1.60 | .28 |
| Average | 1.58 | .07 | 2.11 | .13 | 2.06 | .15 | 2.11 | .16 | 1.80 | .15 |

Note—Confidence was rated from 1 = *guessing* to 3 = *very sure*; auditory detail, feelings and reactions, associations, and trying-to-remember ratings were on a 5-point scale, with 1 = *little or none of that feature* and 5 = *feature clear and distinct*.

suggests that the theme sets that generated the most associations to the lure also led to the most "correct" source attributions, perhaps because such theme sets were the ones most coherently bound as schemas.

Predictive ability of the MCQ and RK measures.

Both the MCQ and the RK questionnaire revealed some differences between true and false memories. On the MCQ, if subjects falsely remembered hearing a critical lure, the memory tended to differ from memories of hearing studied items by its relative lack of auditory detail, lack of information about feelings and reactions to the word, and a lower level of confidence that it was actually heard. For the RK measure, subjects were less likely to assign *remember* ratings to falsely recognized lures than to correctly recognized old items. However, it is also possible that subjects who gave more *remember* ratings or higher ratings of auditory detail and reactions than other subjects also correctly remembered more old items or falsely remembered fewer lures. To investigate such possibilities, individual subjects' RK and MCQ ratings were correlated with their recognition scores (the proportion called old for both lures and old items). The ratings were averaged across responses for lures and old items; when the analyses were done separately for ratings for lures and old items, the same pattern of results was obtained. For the RK measure, the correlations of *remember* ratings both with the proportion of old items correctly recognized ($r = -.04$) and with the proportion of lure items falsely recognized ($r = .00$) were not significant. In short, how often a particular subject used *remember* ratings instead of *know* ratings was not related to their overall performance.

To do a parallel analysis for the MCQ, we used both individual correlations for each characteristic and recognition performance (for both lures and old items) and also multiple regression analysis to assess the overall predictiveness of the MCQ ratings. As can be seen in Table 5, correct recognition performance was significantly correlated with higher ratings of auditory detail, feelings and reactions, and associations on the MCQ. Collectively, the four types of ratings explained 27% of the variance [$F(4,43) = 4.07, p < .01$]. On the other hand, the proportions of false memories were significantly correlated only with higher ratings of associations, and collectively, the ratings explained 18% of the variance [$F(4,43) = 2.30, p < .08$]. This pattern suggests that when contextual detail is more salient in general to a particular person, they will be more likely to correctly recognize old items but not significantly more likely to falsely recognize lures. In contrast, when associations are highly salient to a particular person, that person will tend to recognize more lures as well as more old items.

In summary, both the MCQ and the RK measures indicate that there are differences between true and false memories, although only the MCQ measure accounted for a significant amount of the variance in accuracy from individual to individual.

Table 5
MCQ Ratings as Predictors of Proportion of
Items Correctly Recognized or Lures Called Old

| MCQ Characteristic | <i>r</i> | β |
|---|----------|---------|
| MCQ Ratings as Predictors of Old Items Recognized | | |
| Confidence | .13 | |
| Auditory detail | .39* | .32 |
| Feelings/reactions | .41* | .11 |
| Associations | .36* | .34* |
| Trying to remember | .15 | -.12 |
| $R = .52; R^2 = .27; p < .01$ | | |
| MCQ Ratings as Predictors of Lures Recognized | | |
| Confidence | -.13 | |
| Auditory detail | .14 | .22 |
| Feelings/reactions | .11 | -.22 |
| Associations | .38* | .45* |
| Trying to remember | .21 | -.01 |
| $R = .42; R^2 = .18; p < .08$ | | |

Note—*r* = correlation coefficient; β = standardized regression coefficient. Confidence was not included in the multiple regression analysis because it is considered to be a composite measure related to other measures on the MCQ. * $p < .05$.

DISCUSSION

According to the source-monitoring framework (e.g., Johnson, 1988; Johnson et al., 1993; Johnson & Raye, 1981), both accurate and false memories arise when phenomenal qualities of mental experiences interact with judgment processes. The accuracy of memory depends both on the similarity of information derived from various sources and on the criteria (e.g., perceptual vividness, semantic plausibility, etc.) used to make source attributions. We see these principles illustrated in the present results using the Deese (1959) paradigm, as they have previously been illustrated using other paradigms.

The first question we raised in the introduction was whether true and false memories in the Deese (1959) paradigm have the same qualitative characteristics. This study demonstrated that even when subjects have relatively little contextual information available for veridical memories of old words, the phenomenal characteristics of false memories differed from those of true memories. Memories for studied items were judged as more vivid (i.e., subjects categorized them as *remember* rather than *know*) and were associated with higher levels of remembered auditory detail and information about associated feelings or reactions on the MCQ than were false memories for critical lures. In addition, subjects who tended to give higher ratings for auditory detail and feelings and reactions than other subjects also tended to be more likely to correctly recognize words they had heard (and were not more likely to falsely recognize lures) than were other subjects. Yet despite these differences between the true and false memories, subjects were very likely to falsely recognize the lures. These results suggest that, under some circumstances, salient semantic information can override other information (such as the level of auditory de-

tail) that might be used to increase the accuracy of source monitoring.

Supporting the notion that veridical memories in this paradigm may be associated with higher levels of auditory detail and remembered feelings and reactions is another recent study using MCQ ratings (in a one-speaker blocked format with the Deese paradigm), which found higher ratings for “sound” and “reaction” for recognized studied items than for lures called old (Norman & Schacter, 1997). In addition, a recent positron emission tomography (PET) study found that there was greater blood flow to an area of the brain that processes auditory information (in the left temporoparietal region) when subjects recognized words they had actually heard than when they recognized lures (Schacter et al., 1996).

Our second question concerned whether or not asking subjects about specific characteristics of their memories would induce them to use stricter criteria and thus reduce the rate of false memories (e.g., Dodson & Johnson, 1993; Multhaup, 1995). Although the RK judgments and the MCQ both encourage subjects to think about characteristics of their memories, the MCQ requires a more specific level of examination. In comparison with the RK test condition, the added specificity of the MCQ did not reduce the number of false memories in blocked conditions but did reduce the number of false memories in the random conditions. This suggests that false memories for lures are most vulnerable to additional scrutiny when schemas to structure memory are not as readily available. As proposed in the introduction, blocking the items should increase the chances that the lure will come to mind at study and/or that associated items will come to mind at test to increase inferences that the lure was heard (e.g., Bartlett, 1932; Johnson et al., 1973). That is, associations, like schemas, potentially have effects both at encoding and retrieval (e.g., see Alba & Hasher, 1983, for a review). For example, the likelihood that the lure *needle* would be activated during acquisition in response to *eye* should be greater if *eye* follows *thread* than if it follows an item unrelated to *needle*. Similarly, if the studied item *eye* is activated when the lure *needle* is presented as a test probe, the fact that a list item is rapidly available may be taken as evidence that *needle* had been previously presented and organized with other presented words.

A recent study by Johnson et al. (1997) provides additional evidence that the structure of a test session can affect how true and false memories are processed, as reflected in brain activity. As described earlier, Schacter et al. (1996) found differences between blood flow for veridical memories and blood flow for false memories. As is typical for PET studies, presented test items were grouped by type of item (e.g., 12 lures in a row, then 12 new items in a row). Using an electrophysiological technique in which event-related potentials (ERPs) were recorded, Johnson et al. (1997) compared a grouped test condition with an intermixed test condition (in which the items were not sorted by type). In the grouped condition, the waveforms for old and lure items significantly dif-

fered, whereas they were relatively similar in the intermixed condition. This indicates that how the test list was structured affected how subjects processed the words. Johnson et al. (1997) suggested that subjects were more likely to make judgments on the basis of relative familiarity in the intermixed-test condition and were more likely to access/evaluate specific qualitative characteristics in the grouped-test condition. In a related study, Johnson, Kounios, and Nolde (1996) found different patterns of brain activity when subjects engaged in old/new recognition than when they engaged in source identification for the same items. These brain-imaging studies further support the notion that which memory characteristics are activated depend on the test conditions.

In our study, blocking related items at study also tended to increase the chances that lures (as well as studied items) would be attributed to the speaker who had said the related studied items. These results suggest that under some circumstances, false memories may assume characteristics of related items. Again, as with calling lures old, such an effect of blocking related study items on speaker attribution potentially operates at encoding and at test. At encoding, insofar as subjects rehearsed items in the voices in which they were presented, associatively activated and rehearsed lures would have auditory qualities incorporated into their memory representations (e.g., Johnson, Foley, & Leach, 1988). At test, such speaker information might be directly activated by a test probe (e.g., *needle*). Alternatively, speaker information might be inferred at test on the basis of speaker information that is activated along with related studied items (activated in response to the lure test probe). Subjects may not be able to distinguish the source of the activation of this speaker information (e.g., Henkel & Franklin, in press; see also Reinitz, Lammers, & Cochran, 1992). Thus, at test, false source attributions could come about either as a result of characteristics assumed at encoding or from inferences at test based on properties of other memories.

Regardless of whether the effects obtained here arise primarily at encoding or at test, the results are generally consistent with models that assume that memories are records of neurally distributed cognitive operations that create various attributes (e.g., Johnson, 1992, 1997; Johnson & Hirst, 1993) and that remembering involves attributing whatever characteristics are activated to sources (e.g., Johnson et al., 1993; Johnson & Raye, 1981). In this view, remembering is not the retrieval of a single trace, but rather a process of assembling and evaluating distributed information as a consequence of perceptual and reflective operations. Similarly, other distributed memory models assume that a memory trace is a combination of traces rather than a distinct entity (see Raaijmakers & Shiffrin, 1992, for a review). It is likely that both more conscious inference processes and less consciously controlled activation processes act in the Deese paradigm, combining to create a powerful illusion not only that a lure was heard, but that it was spoken by a particular person.

In summary, as the SMF suggests, veridical and false memories differ from each other in varying degrees, depending upon the context in which they were created (here, blocked or random) and the criteria used to judge whether they were real (here, MCQ or RK). If the context makes it easy to use a schema or associative relations to encode or remember items, then old items that are highly schema relevant will be remembered better. In addition, related items will tend to be falsely recognized (e.g., Anisfeld & Knapp, 1968; Johnson et al., 1973). Especially if an unheard word has had repeated opportunities to be activated (e.g., Underwood, 1965) or is the organizing element of an abstracted schema (e.g., Bransford & Franks, 1971; Posner & Keele, 1968), it will tend to be "remembered." Thus, semantic similarity is particularly powerful in the blocked version of the Deese paradigm because of multiple strong associative relations converging on the critical lures. These compelling associative relations can override perceptual differences (i.e., amount of auditory detail) that might help individuals differentiate true and false memories. The false-memory effect was quite large in the random conditions as well. But in this situation, the influence of semantic similarity evidently can be offset somewhat by a more thorough scrutiny of the memories (MCQ condition), reducing the rate of false memories in comparison with that of veridical memories. More generally, as memories are assembled during the process of remembering, source-monitoring errors may occur when memory records lack sufficient discriminating information or when potentially available source information is ignored in favor of more salient but less differentiating information.

REFERENCES

- ALBA, J. W., & HASHER, L. (1983). Is memory schematic? *Psychological Bulletin*, **93**, 203-231.
- ANISFELD, M., & KNAPP, M. (1968). Association, synonymy, and directionality in false recognition. *Journal of Experimental Psychology*, **77**, 171-179.
- BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- BATCHELDER, W. H., & RIEFER, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, **97**, 548-564.
- BELLI, R., & LOFTUS, E. (1994). Recovered memories of childhood abuse: A source monitoring perspective. In S. J. Lynn & J. W. Rhue (Eds.), *Dissociation: Clinical and theoretical perspectives* (pp. 415-433). New York: Guilford.
- BRANSFORD, J. D., & FRANKS, J. J. (1971). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, **3**, 331-350.
- BRANSFORD, J. D., & JOHNSON, M. K. (1973). Considerations of some problems of comprehension. In W. Chase (Ed.), *Visual information processing* (pp. 383-438). New York: Academic Press.
- CECI, S. J., & BRUCK, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, **113**, 403-439.
- CONWAY, M. A., & DEWHURST, S. A. (1995). Remembering, familiarity, and source monitoring. *Quarterly Journal of Experimental Psychology*, **48A**, 125-140.
- 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., & JOHNSON, M. K. (1996). Some problems with the process dissociation approach to memory. *Journal of Experimental Psychology: General*, **125**, 1-14.
- DURSO, F. T., & JOHNSON, M. K. (1980). The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning & Verbal Behavior*, **19**, 416-429.
- FINKE, R. A., JOHNSON, M. K., & SHYI, G. C. W. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, **16**, 133-137.
- GARDINER, J. M., & JAVA, R. I. (1993). Recognizing and remembering. In A. E. Collins, S. E. Gathercole, M. A. Conway, & P. E. M. Morris (Eds.), *Theories of memory* (pp. 163-188). Hillsdale, NJ: Erlbaum.
- HASHTROUDI, S., JOHNSON, M. K., & CHROSNIAK, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology & Aging*, **5**, 119-126.
- HENKEL, L. A., & FRANKLIN, N. (in press). Reality monitoring of physically similar and conceptually related objects. *Memory & Cognition*.
- HYMAN, I. E., & PENTLAND, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory & Language*, **35**, 101-117.
- JACOBY, L. L., KELLEY, C., BROWN, J., & JASECHKO, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality & Social Psychology*, **56**, 326-338.
- JOHNSON, M. K. (1988). Discriminating the origin of information. In T. F. Oltmanns & B. A. Maher (Eds.), *Delusional beliefs* (pp. 34-65). New York: Wiley.
- JOHNSON, M. K. (1992). MEM: Mechanisms of recollection. *Journal of Cognitive Neuroscience*, **4**, 268-280.
- JOHNSON, M. K. (1997). Identifying the origin of mental experience. In M. S. Myslobodsky (Ed.), *The mythomanias: The nature of deception and self-deception* (pp. 133-180). Mahwah, NJ: Erlbaum.
- JOHNSON, M. K., BRANSFORD, J. D., & SOLOMON, S. K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology*, **98**, 203-205.
- JOHNSON, M. K., FOLEY, M. A., & LEACH, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, **16**, 337-342.
- JOHNSON, M. K., FOLEY, M. A., SUENGAS, A. G., & RAYE, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, **117**, 371-376.
- JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- JOHNSON, M. K., & HIRST, W. (1993). MEM: Memory subsystems as processes. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 241-286). Hillsdale, NJ: Erlbaum.
- JOHNSON, M. K., KOUNIOS, J., & NOLDE, S. F. (1996). Electrophysiological brain activity and memory source monitoring. *NeuroReport*, **7**, 2929-2932.
- JOHNSON, M. K., NOLDE, S. F., & DE LEONARDIS, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory & Language*, **35**, 135-156.
- JOHNSON, M. K., NOLDE, S. F., MATHER, M., KOUNIOS, J., SCHACTER, D. L., & CURRAN, T. (1997). The similarity of brain activity associated with true and false recognition memory depends on test format. *Psychological Science*, **8**, 250-257.
- JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.
- JOHNSON, M. K., RAYE, C. L., FOLEY, H. J., & FOLEY, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, **94**, 37-64.
- LANE, S. M., & ZARAGOZA, M. S. (1995). The recollective experience of cross-modality confusion errors. *Memory & Cognition*, **23**, 607-610.
- LINDSAY, D. S. (1994). Memory source monitoring and eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 27-55). New York: Cambridge University Press.
- LINDSAY, D. S., & JOHNSON, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, **17**, 349-358.

- LINDSAY, D. S., JOHNSON, M. K., & KWON, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, **52**, 297-318.
- LINDSAY, D. S., & READ, J. D. (1994). Psychotherapy and memories for childhood sexual abuse: A cognitive perspective. *Journal of Experimental Child Psychology*, **8**, 281-338.
- LOFTUS, E. F. (1979). The malleability of human memory. *American Scientist*, **67**, 312-320.
- LOFTUS, E. F., DONDERS, K., HOFFMAN, H. G., & SCHOOLER, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition*, **17**, 607-616.
- MCDERMOTT, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language*, **35**, 212-230.
- MULTHAUP, K. S. (1995). Aging, source, and decision criteria: When false fame errors do and do not occur. *Psychology & Aging*, **10**, 492-497.
- NORMAN, K. A., & SCHACTER, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*.
- PAYNE, D. G., ELIE, C. J., BLACKWELL, J. M., & NEUSCHATZ, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory & Language*, **35**, 261-285.
- POSNER, M. I., & KEELE, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.
- RAAJMAKERS, J. G. W., & SHIFFRIN, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology*, **43**, 205-234.
- RAJARAM, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, **21**, 89-102.
- RAYE, C. L., JOHNSON, M. K., & TAYLOR, T. H. (1980). Is there something special about memory for internally generated information? *Memory & Cognition*, **8**, 141-148.
- 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.
- REINITZ, M. T., LAMMERS, W. J., & COCHRAN, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, **20**, 1-11.
- ROEDIGER, H. L., III, & MCDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.
- SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, **2**, 437-442.
- SCHACTER, D. L., REIMAN, E., CURRAN, T., YUN, L. S., BANDY, D., MCDERMOTT, K. B., & ROEDIGER, H. L., III (1996). Neuroanatomical correlates of veridical and illusory recognition memory revealed by PET. *Neuron*, **17**, 267-274.
- SCHOOLER, J. W., GERHARD, D., & LOFTUS, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 171-181.
- SPORER, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology*, **22**, 157-180.
- SPORER, S. L., PENROD, S., READ, D., & CUTLER, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, **118**, 315-327.
- STERN, L. B., & DUNNING, D. (1994). Distinguishing accurate from inaccurate eyewitness identifications: A reality monitoring approach. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 273-299). New York: Cambridge University Press.
- 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). Intrusions of a thematic idea in retention of prose. *Journal of Experimental Psychology*, **103**, 255-262.
- TULVING, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.
- UNDERWOOD, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, **70**, 122-129.
- WEINGARDT, K. R., TOLAND, H. K., & LOFTUS, E. F. (1994). Reports of suggested memories: Do people truly believe them? In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 3-26). New York: Cambridge University Press.
- ZARAGOZA, M. S., & KOSHMIDER, J. W., III (1989). Mised subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 246-255.
- ZARAGOZA, M. S., & LANE, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 934-945.

NOTES

1. Although, using a within-subjects design, it would be interesting to correlate RK and MCQ responses for every item, we wanted to avoid the possibility that judgments in the RK task might be affected by MCQ judgments made for the same item, or vice versa.
2. The full instruction given for associations was, "Indicate to what extent you remember thinking about other things that the word reminded you of when it was said. For example, the word might have made you think of other words on the list or of a topic or other associations you had to the word, or you might remember what came right before or right after it."
3. To check whether the words used as lures were more likely to elicit false alarms even if their related items had not been presented, we compared rates of false recognition for lure items for which no related items had been heard with those for new list items, using a 2 (test: RK, MCQ) \times 2 (item type: new lure, new) mixed-factor ANOVA. There were no significant effects, so in the table we do not distinguish between new lure items and new list items.
4. The Payne et al. (1996) condition was slightly more structured than our two-speaker blocked condition; their speakers alternated saying every other word.
5. For both of these source-monitoring results and the previously described recognition results, consistent findings were obtained when the data were analyzed using a multinomial approach (Batchelder & Riefer, 1990; see Appendix for raw scores used in this analysis).
6. Calculating the R/K responses as a proportion of all responses assigned each type of rating (such that the percentages of R and K responses in a given condition add up to the percent of items correctly recognized as old in that condition) gave the same pattern of results.

APPENDIX
Frequencies of "S," "C," and "N" Responses to
Steven, Cathy, and New Statements by Condition
for Target Items and Critical Lures

| | Source | Response | | | | | |
|---------------------|--------|--------------|-----|-----|----------------|-----|-----|
| | | Target Items | | | Critical Lures | | |
| | | "S" | "C" | "N" | "S" | "C" | "N" |
| MCQ Condition | | | | | | | |
| Blocked one-speaker | S | 170 | 30 | 38 | 41 | 12 | 6 |
| | C | 34 | 161 | 42 | 8 | 39 | 12 |
| | N | 21 | 49 | 226 | 5 | 8 | 46 |
| Blocked two-speaker | S | 104 | 63 | 69 | 21 | 24 | 15 |
| | C | 62 | 117 | 61 | 29 | 21 | 9 |
| | N | 24 | 31 | 242 | 9 | 2 | 46 |
| Random one-speaker | S | 113 | 42 | 82 | 17 | 12 | 29 |
| | C | 45 | 110 | 84 | 9 | 24 | 27 |
| | N | 37 | 48 | 210 | 6 | 11 | 40 |
| Random two-speaker | S | 105 | 82 | 50 | 18 | 15 | 25 |
| | C | 71 | 118 | 51 | 12 | 21 | 27 |
| | N | 42 | 48 | 207 | 3 | 13 | 42 |
| RK Condition | | | | | | | |
| Blocked one-speaker | S | 141 | 33 | 65 | 40 | 7 | 11 |
| | C | 21 | 171 | 48 | 6 | 41 | 13 |
| | N | 24 | 34 | 236 | 5 | 8 | 45 |
| Blocked two-speaker | S | 109 | 63 | 65 | 18 | 25 | 14 |
| | C | 61 | 124 | 55 | 20 | 22 | 17 |
| | N | 19 | 25 | 252 | 6 | 4 | 48 |
| Random one-speaker | S | 109 | 55 | 74 | 19 | 19 | 19 |
| | C | 52 | 111 | 75 | 8 | 26 | 25 |
| | N | 42 | 55 | 199 | 12 | 16 | 31 |
| Random two-speaker | S | 110 | 65 | 62 | 23 | 19 | 14 |
| | C | 68 | 112 | 58 | 21 | 23 | 14 |
| | N | 42 | 55 | 200 | 11 | 10 | 38 |

(Manuscript received June 11, 1996;
revision accepted for publication October 4, 1996.)