Journal of Experimental Psychology:
Learning, Memory, and Cognition
1998, Vol. 24, No. 4, 1026–1040

0278-7393/98/$3.00

The Phenomenology of False Memories: Episodic Content and Confidence

Jonathan B. Holmes, Harriet Salatas Waters, and Suparna Rajaram
State University of New York at Stony Brook

Three experiments investigated the phenomenology of false memories in J. D. Bransford and J. J. Franks's (1971) semantic integration paradigm using Remember–Know judgments and confidence ratings. Findings included high rates of Remember false alarms that increased with greater integrative demands, a dissociation between Remember and Know judgments in the standard Bransford and Franks condition, and similar patterns of Remember and Know judgments in a condition where semantic integrative processes were weakened. Confidence ratings were consistently higher for Remember judgments except where integrative processes were greatest. These results are discussed in terms of the phenomenology of true and false memory for episodes.

Both the study of consciousness and phenomenology and the study of false memories are currently experiencing a great resurgence in psychology. The former, consciousness and phenomenology, have gone from the primary facts of mental life to be explained by psychology to nonquestions in the light of verificationist methodology and back again to central issues. This resurgence can be seen in such areas as the study of conscious versus nonconscious processing (for reviews, see Holender, 1986; Shanks & St. John, 1994; Velmans, 1991), implicit versus explicit processes (Reber, 1991; for review, see Schacter, 1987), amnesia (for reviews, see Moskovitch, Vriezen, & Gottstein, 1993; Shimamura, 1986), and a large volume of writing concerning consciousness itself (e.g., Baars, 1988; Dennett, 1991; Searle, 1992). The latter, false memory, is now both a scientific and popular interest, due in part to current clinical and legal concerns surrounding recovered memories of childhood trauma (e.g., Pope, 1996; Schacter, 1996; Schooler, 1994), as well as to renewed interest in memory illusions (for review, see Roediger, 1996). Finally, the areas of phenomenology and memory have intersected, as in work on the phenomenology of recollective experience (Gardiner, 1988; Tulving, 1985; for reviews, see Gardiner & Java, 1993; Rajaram & Roediger, 1997), explorations of phenomenological cues used in reality and source monitoring (Johnson, 1988; Johnson & Raye, 1981; for review, see Johnson, Hashtroudi, & Lindsay, 1993), and work on memory errors from misinformation effects (Loftus, 1979, 1992; Zaragoza & Lane, 1994).

The phenomenology of veridical memory has also seen greater attention since Tulving's (1985) introduction of *Remember* and *Know* judgments as measures of conscious experience accompanying memories (e.g., Gardiner, 1988; Gardiner & Java, 1990; Gardiner & Parkin, 1990; Rajaram, 1993, 1996). However, until very recently not much attention has been directed toward applying Remember and Know judgments to false memories (e.g., Conway, Collins, Gathercole, & Anderson, 1996; Lane & Zaragoza, 1995; Roediger & McDermott, 1995). Tulving (1985, p. 1) noted that "One might think that memory should have something to do with remembering, and remembering *is* a conscious experience." The observation is just as pertinent for false memories, as having a false memory is also a conscious experience. Moreover, only through concurrent study of false memory can one hope to gain full insight into the nature of memory as a whole, both true and false. In this article we addressed the phenomenology of memories arising from semantic integration, a particular form of memory reconstruction.

Previous Research on False Memories

Early in psychology's history, memory research was the study of conscious, phenomenal memory (Angell, 1909; Strong, 1913; for review, see Brewer, 1992). A significant gap in attention to false memory phenomenology soon occurred, however, and lasted until around the 1970s. Since then, phenomenological experience has slowly regained an important position in memory research, particularly in recent years.

One area of inquiry concerns the phenomenology of false memories arising from misinformation (Loftus, 1979, 1992). In a typical misinformation experiment, participants might

Jonathan B. Holmes, Harriet Salatas Waters, and Suparna Rajaram, Department of Psychology, State University of New York at Stony Brook.

This article was in partial fulfillment of Jonathan B. Holmes's PhD requirements in cognitive psychology at the State University of New York at Stony Brook. Portions of this article were previously presented in 1996 at the 36th Annual Meeting of the Psychonomic Society, Los Angeles; in 1995 at the 7th Annual Convention of the American Psychological Society, New York; and in 1995 at the 66th Annual Meeting of the Eastern Psychological Association, Boston. We thank our reviewers for their generous and helpful comments during the preparation of this article.

Correspondence concerning this article should be sent to Jonathan B. Holmes, c/o Harriet Salatas Waters, Department of Psychology, State University of New York, Stony Brook, New York 11794. Electronic mail may be sent to JHolmes@ccvm.sunysb.edu.

watch a videotape of a simulated crime. They then might be asked, "Did the thief who was tall and wearing a hat leave the window up after he left the house?" In the video, the thief had not worn a hat, and this type of subtle, postevent misinformation produces false memories, such as believing that the thief actually had worn a hat. Moreover, participants often rate their false memories with high degrees of confidence, indicating a firm belief in their false memories. Indeed, the distortions that can occur appear to be almost endless, including "seeing nonexistent items, such as broken glass, tape recorders, and even something as large and conspicuous as a barn (in a bucolic scene that contained no buildings at all)" (Loftus, 1992, p. 121).

Another study applying confidence ratings to false memories was not actually a typical false memory study but rather an investigation of the effects of memory for verbal materials. Bransford and Franks (1971) found a large false memory effect arising from what they termed *semantic integration.* Participants were auditorily presented with a number of sentences of varying length that could be put together to form complete ideas. For example, participants might be presented with the sentences *The jelly was sweet, The ants were in the kitchen, The jelly was in the kitchen, The jelly was on the table,* and *The ants ate the sweet jelly on the table.* Following each sentence was a short semantic orienting question, such as *Did what?,* to which participants would respond *Ate the sweet jelly on the table.* Following a retention interval, participants were presented with a set of similarly constructed sentences, half that had been previously presented and half that had not. They were then asked how confident they were that they had or had not previously encountered each sentence.

There were two principle results. First, participants essentially could not discriminate between old and new sentences in either recognition performance or confidence ratings. Second, as sentence complexity increased, confidence levels that the sentence had been previously presented increased regardless of whether the sentence had actually been previously encountered. For example, if two recognition sentences were *The ants were in the kitchen* and *The ants in the kitchen ate the sweet jelly,* the latter more complex sentence would be rated with higher confidence whether or not it had been previously encountered at study. This was termed the *linear abstraction effect,* as confidence increased in a linear fashion with increasing sentence complexity. These two effects, poor discrimination and linear abstraction, were seen as the products of one underlying memory process—the creation of an integrated gist representation. Poor discrimination occurred because the verbatim sentence structure was never represented, and linear abstraction effects occurred because longer sentences expressed more of a complete idea, therefore more closely matching the integrated gist that participants were presumably remembering.

Complementing such research using confidence ratings are two areas focusing on other aspects of phenomenology. One area of research involves finding what phenomenological cues people use during reality monitoring, that is, distinguishing between events they actually experienced and events they have internally generated through imagination (Johnson, 1988; Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981). Reality monitoring theory states that memories for actual events should on average rate higher on certain phenomenological measures, specifically sensory–perceptual detail, temporal–spatial detail, semantic detail, and affect (Johnson et al., 1993). Memories for imagined events should on average rate higher on apperceptive processes or knowledge of cognitive operations.

This proposal turns out to be true. In one study participants were asked to recall either actual events or dreams and unfulfilled fantasies. Memories for actual events scored higher on visual detail, sound, smell, realism, spatial and temporal context, and supporting memories (Johnson, Foley, Suengas, & Raye, 1988). In contrast, memories for internally generated events were rated higher on complexity, implications, and intensity and were thought about more often. More important, these same phenomenological characteristics are also used by people to assess the source of their memories (Johnson et al., 1988; Suengas & Johnson, 1988).

However, the distribution of characteristics for true and false memories does have some degree of overlap, and this overlap becomes more problematic under certain conditions. For instance, people may fail to correctly use reality monitoring cues and therefore confuse the source of their memories (e.g., Schooler, Gerhard, & Loftus, 1986). As well as simply failing to attend to reality monitoring cues, certain experimental conditions will lead to poorer reality monitoring. For example, increasing the semantic similarity between information from different sources makes it more likely to confuse the sources (Johnson, Raye, Foley, & Foley, 1981). Finally, identifying the source of one's memory may not be so much a process of calling up memory for a source as a strategic process at retrieval aimed at evaluating the most likely source of a memory (Johnson et al., 1993). For instance, given certain information one knows, certain sources may be ruled out a priori without any real assistance from actual memory processes.

A second area of research assessing memory phenomenology beyond confidence comes from applying Remember and Know judgments to measure certain phenomenological characteristics of memories (Gardiner, 1988; Tulving, 1985). In a typical Remember–Know experiment, participants are presented with a number of items. After a variable retention interval, they receive a recall or recognition test. Accompanying their recall or recognition judgments, they are asked to make Remember or Know judgments on their memories. A Remember judgment indicates that a person had a conscious recollection of having encountered an item at study accompanied by supporting memories. A Know judgment indicates that they know the item was from study but without the distinct conscious recollection of having encountered the item at study and without associated supporting memories.[1]

[1] Originally, Tulving (1985) conceived of Remember judgments as reflecting the operation of the episodic memory system, whereas Know judgments reflected the operation of the semantic or procedural system. Subsequent to Tulving's original work, the question of what memory processes lie behind Remember–Know

Remember and Know judgments have very recently been applied to false as well as true memories, with the general and somewhat surprising finding that false memories can be rated as Remember judgments (Lane & Zaragoza, 1995; Roediger & McDermott, 1995; for review, see Roediger, 1996). In one study addressing cross-modality confusion errors, participants were presented with a mixed list of pictures and words and then asked whether they had encountered an item at study as a picture (Lane & Zaragoza, 1995). False alarms from cross-modality confusions were often rated as Remember judgments. Lane and Zaragoza (1995) noted that, "neither remembering or knowing was uniquely diagnostic with regard to the actual truth of a memory" (p. 609).

Roediger and McDermott (1995) recently applied Remember–Know judgments to a simple list-learning paradigm modeled after Deese (1959). Deese showed that when people were presented with a list of primary associates to some word, such as *needle,* they would recall having encountered the primary associate even though it had not been on the list. Roediger and McDermott replicated Deese's methods and added confidence and Remember–Know judgments. The results showed people to be highly confident in their false alarms (Experiment 1), and to often rate false alarms as Remember judgments (Experiment 2; see also Read's 1996 work with the Deese paradigm).

In sum, false memories appear be able to share phenomenological characteristics with true memories. They can be rated with Remember judgments, indicating a conscious awareness and supporting memories for an event that never occurred (Lane & Zaragoza, 1995; Roediger & McDermott, 1995). As well, false memories can be rated with high confidence, indicating a firm belief that the memory is actually true (Bransford & Franks, 1971; Loftus, 1979, 1992; Roediger & McDermott, 1995). One interesting question is how well reality-monitoring theory fits into a picture of false memory where false memories can share such dramatic phenomenology with true memories. We address this issue more in the General Discussion; however, for the moment, suffice it to say that from one point of view reality monitoring does fit nicely with this picture. A firm belief in a false memory would seem more likely if it did have the phenomenological characteristics one would expect a true memory to have.

The Present Study

In the three experiments that follow, we investigated the phenomenology of false memories using Remember–Know judgments and confidence ratings. To reliably generate a significant number of false memories, we selected the Bransford and Franks (1971) semantic integration paradigm. There were several reasons for this choice. First, Bransford and Franks's findings are classic, well-documented effects (e.g., Cofer, 1973; Griggs, 1974; Peterson & McIntyre, 1973; Singer, 1973). Second, in light of the relatively new Remember–Know measures, the phenomenology of the Bransford and Franks effect can be more fully explored. Third, semantic integration occurs every day; rarely do we remember the verbatim structure of what we read or hear. The real-world character of this particular type of memory process should lead us away from more artificial tasks toward a true view of everyday false memory phenomenology. Finally, the Bransford and Franks paradigm presents a dynamic look at false memories. Not only is there almost no discrimination between old and new sentences, but confidence ratings increase over increasing sentence complexity (whether a sentence was presented or not). In sum, we had an opportunity to investigate the relationship between the phenomenology represented by Remember judgments and false memories as confidence levels changed systematically with increased integration.

Experiment 1

Our aim in Experiment 1 was to replicate the Bransford and Franks (1971) false memory effect arising from integrative processes and to assess the resulting recognition with Remember–Know judgments. Remember–Know research concerned with veridical memory does not produce many false alarms, and what false alarms are found are normally rated with Know judgments (Gardiner, 1988; Gardiner & Java, 1990; Rajaram, 1993). However, similar research concerned with false memory finds false memories often rated with Remember judgments (Lane & Zaragoza, 1995; Roediger & McDermott, 1995).

Tulving (1985) originally conceived of Remember judgments as measuring awareness associated with the workings of the episodic memory system. We view Remember judgments as measuring a subset of what we term *episodic content.* Episodic content refers to the content of a memory that makes it feel like an event from one's personal past, an episode one has experienced and taken in, which can be reproduced in memory. Not all episodic content is encompassed by Remember judgments, only that part that concerns awareness of having experienced an event with associated

judgments has become more complicated. As Johnson (1988, p. 392) noted, "Although the distinction between remembering and knowing captures a clear phenomenal difference between mental experiences, evidence for a meaningful theoretical difference has been harder to come by." For instance, it remains possible that Remember judgments reflect the functioning of the episodic memory system, and Know judgments reflect either semantic or procedural systems. On the other hand, in the vein of two-process theories of memory, Remember judgments may reflect more explicit, conceptual processing, whereas Know judgments may reflect processing based on familiarity, perceptual fluency, and more implicit processing (for reviews, see Gardiner & Java, 1993; Roediger, Wheeler, & Rajaram, 1993; Rajaram & Roediger, 1997). Further complicating the issues is the question of whether whatever processes drive Remember and Know judgments are mutually exclusive, independent of each other, or some mix (Yonelinas & Jacoby, 1995). These are important issues for discovering the nature of memory processes, although not particularly pertinent to questions of phenomenology. Whatever the underlying processes, Remember judgments *do* capture the phenomenological flavor of an episode, perhaps even an autobiographical experience, from one's past.

memories, which may involve sensory and contextual details. Moreover, though the term *episodic content* may seem to be the defining feature of episodic memory, we are only using the term to refer to *phenomenology,* rather than to any specific type of memory system.

The fact that false memories can be rated with Remember judgments indicates that false memories can have episodic content.[2] In the original Bransford and Franks experiment (1971), people reported high levels of confidence for their false memories, especially when integrative processes were strongest. If they actually believed their false memories to be true, then these false memories should share phenomenological characteristics normally associated with true memories (i.e., episodic content). At the same time, reality monitoring research has shown that sensory and contextual details, as well as supporting memories, are on average more closely associated with memories for actually occurring events (Johnson et al., 1988). We proposed to both accommodate and extend these findings by demonstrating that episodic content is not necessarily more closely associated with veridical memory, particularly in contexts in which there is substantial semantic integration. Because Remember judgments entail sensory and contextual details, as well as supporting memories, we predicted that false memories produced by integrative processes would be rated with Remember judgments, representing the presence of episodic content, the very phenomenology a memory should have if it was believed to be true.

Method

Participants. Forty undergraduates served as voluntary participants, receiving experimental credit toward psychology class requirements for their participation.

Materials. The materials included an original list of 48 sentences, four acquisition lists of 24 sentences from the original list, four recognition lists of all 48 sentences from the original list, and a set of Remember–Know instructions. All sentence list construction followed the methods used by Bransford and Franks (1971, Experiment 3).

Construction of the original list of 48 sentences began by taking 4 original sentences from Bransford and Franks (1971), each expressing one coherent idea that could be divided into four smaller "idea units." For each complex sentence (4s), 11 smaller sentences were created from either single idea units or combinations of idea units, each smaller sentence expressing some portion of the entire semantic content of the complex sentence. For example, the four-idea-unit sentence *The scared cat running from the barking dog jumped on the table* could be broken down into the three-idea-unit sentence *The cat running from the barking dog jumped on the table,* the two-idea-unit sentence *The cat was running from the barking dog,* and the one-idea-unit sentence *The cat was running from the dog.* For each four-idea-unit sentence, there were 3 three-idea-unit sentences (3s), 4 two-idea-unit sentences (2s), and 4 one-idea-unit sentences (1s) for a total of 12 sentences for each of the 4 original sentences.

Four acquisition lists were then created from the original list of 48 sentences. Following Bransford and Franks (1971), the original list was split in half to create Acquisition Lists 1 and 2, so that the sentences on List 1 were not presented in List 2, and vice versa, though making sure each idea set was completely expressed on each list. The sentences in List 1 were randomized and sorted two separate times, creating Acquisition Lists 1a and 1b (similarly for sentences in List 2 to create Lists 2a and 2b). Acquisition Lists 1a and 1b contained the same sentences, but in different orders (as for Lists 2a and 2b), whereas Lists 1a and 1b contained different sentences than Lists 2a and 2b. Each acquisition list was structured so (a) the order of sentences and the number of idea units in each sentence was random, (b) each idea set was represented once in a block of four sentences, and (c) no sentence from an idea set followed another sentence from that same idea set. Each acquisition list sentence was paired with an elliptical question asking about some semantic aspect of that sentence. For example, the sentence *The scared cat was running from the barking dog* might be followed by the elliptical question *Which cat?* (to which participants would hopefully reply, *The scared cat.*). Each constituent of each idea was questioned approximately as often as every other constituent. The number of idea units in a sentence was counterbalanced according to Bransford and Franks (1971).

Four recognition lists were then created. Each recognition list contained all 48 sentences from the original list, though was structured similarly to the acquisition lists with respect to randomization and blocking rules. This meant each recognition list contained half old sentences from the study and half new sentences not from the study. Each recognition list was used equally often with participants who had received either Acquisition List 1 or 2.

Acquisition and recognition lists were recorded in a male voice on a Sony portable stereo tape player. Each acquisition list sentence was followed approximately 4 s later by an elliptical question. Each elliptical question was followed approximately 10 s later by the next acquisition sentence.

Procedure. Groups of 1 to 4 participants were tested at a time. Participants were first randomly assigned to either List 1 or List 2. Then they were randomly assigned to either Acquisition Sublist a or b of the assigned list, and Recognition Sublist a or b of the assigned list.

Participants listened to an acquisition set of 24 sentences and wrote answers to each elliptical question on an answer sheet, following the methods of Bransford and Franks (1971). Participants then worked on two filler tasks, together taking 15 min. First, the experimenter asked participants to list all the towns, cities, or states anywhere in the world. After approximately 10 min, the experimenter then asked participants to list as many U.S. presidents as they could think of. After the filler tasks, participants read recognition and Remember–Know instructions silently to themselves as the experimenter read them aloud. The recognition and Remember–Know instructions followed closely those of Gardiner (1988) except for necessary changes. One participant of the group was then asked to explain the difference between Remember and Know judgments to the experimenter. The experimenter attempted to clarify or correct any misunderstandings or uncertainties concerning the recognition and Remember–Know instructions for the group. Filler tasks and Remember–Know instructions together provided for a 20-min retention interval between study and test. After the experimenter was certain the participants understood the nature of Remember–Know judgments, participants then listened to a set of 48 recognition sentences one at a time and were given

[2] It may seem odd to postulate that "false episodic content" could actually exist, as almost by definition, episodic content requires an episode to occur. However, when we say "false episodic content" we are referring to a memory phenomenology which makes a memory seem like an episode from one's personal past. We are not specifying whether false episodic content is necessarily the same phenomenon as true episodic content. The two simply share phenomenology.

unlimited time to make recognition and Remember–Know judgments for each sentence on an answer sheet before proceeding to the next recognition sentence.

Participants were not informed that there would be a recognition and Remember–Know judgment phase to the experiment at the time of study, nor were they informed that the middle tasks were filler tasks. Full debriefing of the nature and purpose of the experiment occurred at the end of the experiment. The entire experimental session took approximately 50 min.

Results

Presented below are analyses for overall recognition and Remember–Know judgments. All analyses were computed on mean proportioned frequencies of total "yes" responses to control for the unequal occurrences of sentences of differing numbers of idea units. The recognition analysis compared overall hits and false alarms separately across the number of idea units with one-way repeated measures analyses of variance (ANOVAs) to check for overall linear abstraction effects. Where applicable, post hoc Newman–Keuls tests were completed to evaluate possible differences between individual idea units. This was followed by dependent measures t tests between the same number of idea unit points to determine whether participants were discriminating between old and new sentences. Similar statistical procedures were then used for Remember hits and false alarms and Know hits and false alarms.

Recognition. Results for overall recognition are shown in Figure 1. There were overall linear abstraction effects (increasing rates over sentence complexity) for both hits and false alarms, $F(3, 117) = 44.57$, $MSE = 0.011$, $p < .001$, and $F(3, 117) = 60.48$, $MSE = 0.011$, $p < .001$, respectively. Post hoc Newman–Keuls tests for hits showed all idea units differing significantly ($ps < .01$). Newman–Keuls tests for false alarms showed all idea units differing at the .01 level except for 3s versus 4s ($p = .09$). T-test comparisons between hits and false alarms at each idea level showed participants discriminating significantly between old and new sentences for 1s, 2s, and 4s ($ps < .05$), though with only marginal significance for 3s, $t(39) = 1.98$, $p = .06$. However, though participants were discriminating between old and new sentences, it can be seen from Figure 1 that whatever discrimination occurred was rather minor. In sum, the results mirror those of Bransford and Franks (1971), with linear abstraction effects for both hits and false alarms. Our results showed slightly increased discrimination between old and new sentences than Bransford and Franks originally found, though this is actually more in keeping with subsequent replications (Griggs, 1974; Peterson & McIntyre, 1973; Singer, 1973). There remain, however, relatively high rates of false alarms.

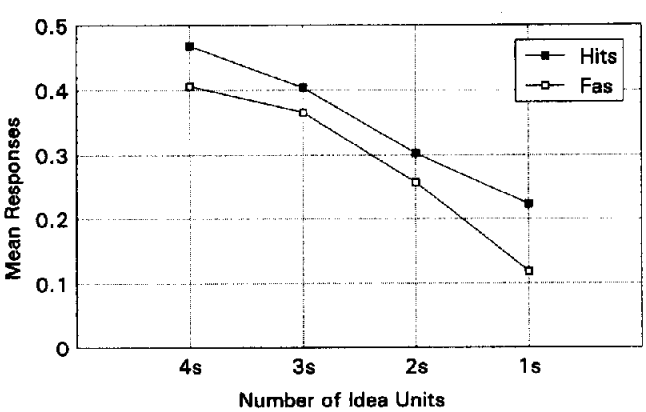

Figure 1. Mean rates of overall recognition hits and false alarms (Fas), Experiment 1.

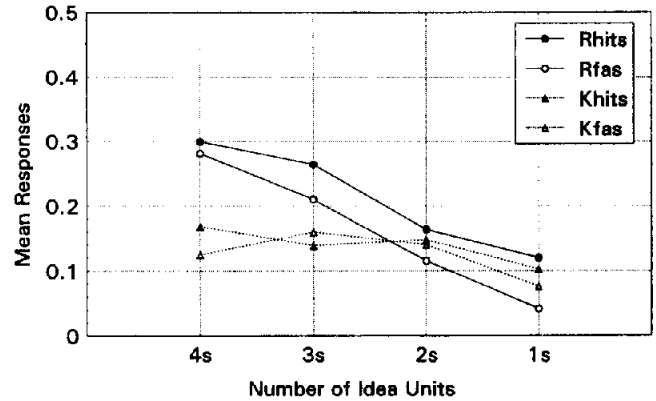

Figure 2. Mean rates of Remember–Know hits and false alarms, Experiment 1. Rhits = Remember hits; Rfas = Remember false alarms; Khits = Know hits; Kfas = Know false alarms.

Remember–Know judgments. Results for Remember–Know judgments are shown in Figure 2. For Remember judgments, as with overall recognition, there were linear abstraction effects for both hits and false alarms, $F(3, 117) = 16.57$, $MSE = 0.017$, $p < .001$, and $F(3, 117) = 40.55$, $MSE = 0.011$, $p < .001$, respectively. Post hoc Newman–Keuls tests for Remember hits showed all idea units differing from one another ($ps < .01$) except for 1s versus 2s and 3s versus 4s. Newman–Keuls tests for Remember false alarms showed all idea units differing from one another ($ps < .01$). Pairwise t tests showed participants discriminating between old and new sentences for 1s, 2s, and 3s ($ps < .05$), though not for 4s ($p = .63$). These results show Remember judgments mirroring the pattern seen in overall recognition, with some discrimination and, more interestingly, a linear abstraction effect for both hits and false alarms.

For Know judgments, there was no linear abstraction effect for hits, though there was for false alarms, $F(3, 117) = 4.63$, $MSE = 0.011$, $p < .01$. Nonetheless, mean differences across idea units did not appear to show a consistently increasing pattern of Know false alarms for more complex sentences. In keeping with this impression, post hoc Newman–Keuls tests for false alarms showed only 1s differing significantly from 2s, 3s, and 4s ($ps < .05$), which did not differ from one to another. Pairwise t tests showed no significant discrimination between old and new sentences for Know judgments at any idea unit level. These results show Know judgments displaying a very different pattern of results than seen in overall recognition and Remember

judgments. There was not a significant linear abstraction effect for Know hits, and a substantially reduced effect for Know false alarms (only 1s differed from other idea units). As well, there was no significant discrimination between old and new sentences for Know judgments, whereas overall recognition and Remember judgments did show significant, though minor, levels of discrimination.

Discussion

The current recognition results replicated Bransford and Franks's (1971) linear abstraction effects for overall recognition and poor discrimination between old and new sentences. Bransford and Franks attributed the poor discrimination to participants abstracting the "gist" rather than verbatim sentence structure. They attributed the linear abstraction effect as due to more complex sentences better matching the gist information on which participants were judging their recognitions.

The main purpose of Experiment 1, however, was to investigate the phenomenology behind the memories in Bransford and Franks's (1971) study. Remember judgments mirrored the results for overall recognition, with poor discrimination and linear abstraction effects for both hits and false alarms. On the other hand, Know judgments showed a flattening of the effect, that is, no linear abstraction effects, and no discrimination between old and new sentences. These results suggest two things. First, it was more likely for a memory to be rated with a Remember judgment when integrative processes were strongest. Know judgments appeared sensitive to integrative processes only insofar as they showed no discrimination between old and new sentences. In keeping with the notion that Remember judgments measure episodic content, the results suggest that the majority of memories underlying Bransford and Franks's linear abstraction effect had episodic content. Second, there were large numbers of Remember false alarms, indicating that participants were having a conscious awareness of encountering sentences, with associated memories, which were never presented. False memories from semantic integration were actually more likely than not to have episodic content. These latter results are in general agreement with other recent studies of false memories showing false memories rated as Remember judgments (Lane & Zaragoza, 1995; Roediger & McDermott, 1995). Specifically, they nicely compliment the findings of Roediger and McDermott, who found high levels of Remember false alarms for semantically based intrusions.

The present results indicate that participants were reporting phenomenological characteristics typical of real or true memories for memories of experiences they never had. How this transference of phenomenological characteristics of real memories to false memories is accomplished is the key question under investigation. We propose that semantic integration processes not only increase participants' confidence in false memories as Bransford and Franks (1971) reported but also encourage the construction of episodic content for these false memories. Furthermore, we propose that this occurs by means of a two-step process in which semantic integration processes produce not only an integrated gist representation but also allow for the transference of episodic content from individual sentences to this gist representation. The presence of episodic content then drives up a person's confidence in real and false memories alike, with its perceived presence more likely the more closely a sentence maps onto the integrated gist representation. Whether a closer mapping of sentence onto gist representation is an independent component of confidence (the sole component that Bransford and Franks identified), in addition to the increased likelihood of episodic content with more complex sentences, remains an empirical question. Thus, the proposed role of episodic content in confidence judgments was the focus of Experiment 2.

Experiment 2

In Experiment 1, we replicated the Bransford and Franks (1971) effect and showed a dissociation between Remember–Know judgments such that Remember judgments closely mirrored overall recognition, whereas Know judgments showed a different pattern of results indicating less sensitivity to semantic integrative processes. Using a 3-point scale, Tulving (1985) originally showed that Remember judgments were accompanied by generally high levels of confidence, whereas confidence for Know judgments varied between medium and low. Subsequent work comparing levels of confidence for Remember–Know judgments used a dichotomous confidence scale of *sure* versus *unsure* and showed that confidence did not always mirror the results from Remember–Know judgments (Gardiner & Java, 1990; Rajaram, 1993). In Roediger and McDermott's (1995) recent investigation of false memory, false recognitions were rated with high confidence in Experiment 1 (using a 4-point scale) and often labeled with Remember judgments in Experiment 2. Unfortunately, Remember–Know judgments and confidence ratings were not assessed in the same experimental session. In Experiment 2 of the present investigation, we wished to both replicate the results from Experiment 1 and further explore the relationship between Remember–Know phenomenology and confidence using the Bransford and Franks's (1971) paradigm, all in a single session.

To this end, we constructed a 5-point confidence scale to more reliably assess the relationship between confidence and Remember–Know judgments. As well, because our primary goal was to make direct comparisons between confidence and Remember–Know judgments, participants were asked to make confidence judgments directly subsequent to making recognition and Remember–Know judgments. Although it could be argued that multiple judgments might affect performance in unforeseen ways, our view was that as long as patterns of overall recognition and Remember–Know judgments paralleled findings from Experiment 1, that possibility could be ruled out in making a direct comparison between Remember–Know judgments and confidence in the current experiment.

As indicated in the Experiment 1 *Discussion*, the semantic integration hypothesis we have proposed to account for different patterns of Remember–Know judgments would

predict both higher confidence for Remember compared with Know judgments and higher confidence with increases in sentence complexity, at least for Remember judgments. Any predictions concerning Know judgments must be tempered by the fact that there is no inherent contradiction in being certain an item has been presented, though no distinct phenomenology beyond familiarity and no supporting memories arise (Strack & Förster, 1995). Nevertheless, though Know judgments could be rated with high confidence, in the current experimental context it seemed unlikely. Of greater concern was whether confidence scores for Know judgments would show sentence complexity effects as predicted for Remember judgments. If the presence of episodic content is the sole driving force behind confidence, the answer would be no. If, as Bransford and Franks (1971) had originally suggested, a closer match between sentence and gist representation increases confidence, then some sentence complexity effects might be present, even without reported episodic content, that is, with Know judgments.

Method

Participants. Forty undergraduates served as voluntary participants, receiving experimental credit toward psychology class requirements for their participation.

Materials and procedure. The materials and procedure for Experiment 2 were identical to those from Experiment 1, with the following exception: Participants were requested to also make confidence judgments for each test sentence after making their recognition and Remember–Know judgments. We constructed a 5-point confidence scale (5 = *very confident*, 1 = *not confident*). We chose a 5-point scale for two reasons. First, Bransford and Franks (1971) originally used a 5-point scale in their study, and we followed this precedent. Second, confidence scales in Remember–Know studies have usually been on only a 2-point scale, with the exception of Tulving's (1985) 3-point scale. We wanted to use a more sensitive measure and therefore applied a wider confidence scale than in previous Remember–Know studies. Recognition instructions for making Remember–Know judgments and confidence explicitly asked participants to (a) state whether the sentence had been presented or not; (b) if yes, to make Remember–Know judgments; and (c) to state how confident they were that the sentence had been presented (not how confident they were in their Remember–Know judgments).

Results

The same strategy for analyses was used in Experiment 2 as in Experiment 1. Means for overall recognition are shown in Figure 3. As in Experiment 1, there were linear abstraction effects for hits and false alarms, $F(3, 117) = 39.17$, $MSE = 0.010$, $p < .001$, and $F(3, 117) = 65.16$, $MSE = 0.010$, $p < .001$, respectively. Post hoc Newman–Keuls tests showed all idea units differing from each other for hits and false alarms, except for 3s versus 4s for hits ($p = .19$). Pairwise t tests showed participants discriminating between old and new sentences for 1s, 2s, and 3s ($ps < .001$), though not for 4s ($p = .53$). Overall, the results for overall recognition replicate those from Experiment 1, showing some discrimination between old and new sentences and clear-cut linear abstraction effects.

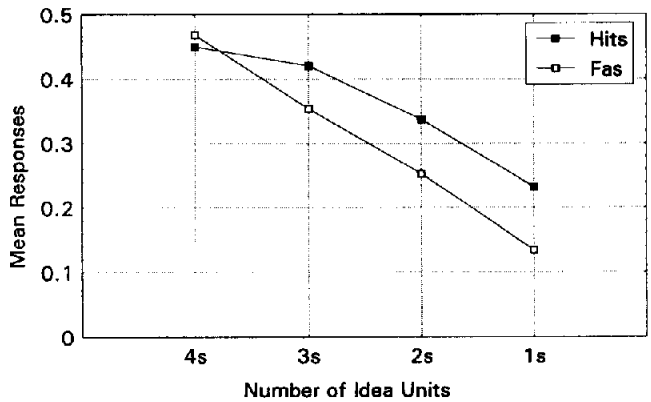

Figure 3. Mean rates of overall recognition hits and false alarms (Fas), Experiment 2.

Remember–Know judgments. Results for Remember–Know judgments are shown in Figure 4. For Remember judgments there were linear abstraction effects for both hits and false alarms, $F(3, 117) = 23.99$, $MSE = 0.015$, $p < .001$, and $F(3, 117) = 26.65$, $MSE = 0.013$, $p < .001$, respectively. Post hoc Newman–Keuls tests showed all idea units differing from each other for hits and false alarms, except for 3s versus 4s for hits ($p = .13$). Pairwise t tests showed participants significantly discriminating between old and new sentences, though at low levels, for 1s, 2s, and 3s ($ps < .05$), though not for 4s ($p = .16$). The results for Remember judgments mirror those for overall recognition, again repeating the results from Experiment 1.

For Know judgments, there was no linear abstraction effect for hits, though there was a significant effect for false alarms, $F(3, 117) = 5.44$, $MSE = 0.011$, $p < .01$. Post hoc Newman–Keuls tests for false alarms showed that this effect was due only to 1s differing from 2s, 3s, and 4s ($ps < .05$), with no differences between these latter means. Pairwise t tests showed participants significantly discriminating between old and new sentences only for 1s, $t(39) = 3.44$, $p < .001$. Again, the overall Know judgments results mirrored

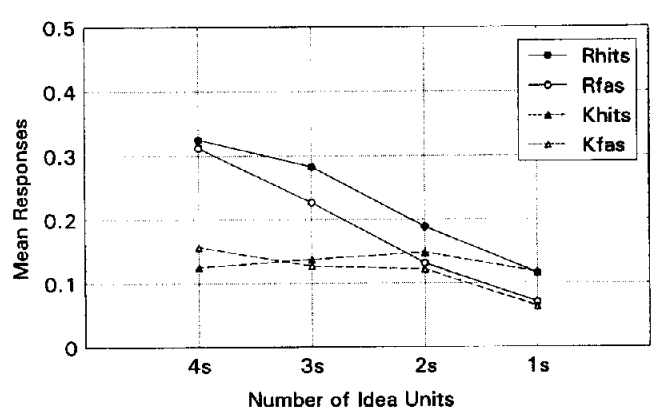

Figure 4. Mean rates of Remember–Know hits and false alarms, Experiment 2. Rhits = Remember hits; Rfas = Remember false alarms; Khits = Know hits; Kfas = Know false alarms.

those from Experiment 1, with no discrimination between old and new sentences (except slightly for 1s in Experiment 2), and attenuated at best linear abstraction effects (nonsignificant for hits, and only 1s differing from other idea units for false alarms).

Confidence ratings. With the results from Experiment 1 replicated in Experiment 2, the stage was set for an examination of the relationship of confidence ratings to Remember–Know judgments across levels of complexity, the main purpose of Experiment 2. Results of confidence judgments are shown in Figure 5. Because participants did not necessarily have hits or false alarms for all sentence complexity levels for Remember and Know judgments, there was a certain amount of missing data in the confidence analyses. To include as many participants as possible in each one-way ANOVA assessing linear abstraction effects, we collapsed two of the four levels along the sentence complexity dimension. For Remember hit and false alarm ratings, scores for sentences with one idea unit were combined with sentences with two idea units because a majority of participants who had only three out of the four data points had missing data on the low end of the dimension (62%). A similar approach was taken for Know hits and false alarms, except this time scores for three- and four-idea-unit sentences were combined. In contrast to the Remember data, participants with missing Know data usually had their missing data point on the other end of the sentence complexity dimension (81% of the time). This difference in proportions of missing data on either end of the sentence complexity dimension was, in fact, significant, as shown by a z test on proportions ($z = 3.00$, $p < .01$, two-tailed).

Confidence ratings for Remember judgments showed a linear abstraction effect for hits ($n = 28$), $F(2, 54) = 5.22$, $MSE = 0.246$, $p < .01$, and for false alarms ($n = 27$), $F(2, 52) = 13.24$, $MSE = 0.157$, $p < .001$.[3] Post hoc Newman–Keuls tests for the Remember hit data indicated that confidence for three- and four-idea-unit sentences did not differ significantly, but both differed from the confidence rating given the ½ idea unit sentences ($p < .01$). Newman–Keuls tests on the false alarm data indicated that 3 and 4 idea unit sentences differed at the .10 level, and both differed significantly from ½ idea unit sentences ($p < .01$). Pairwise t tests conducted at each level of sentence complexity did not show any discrimination between confidence ratings for Remember hits versus Remember false alarms. Thus, confidence ratings applied to Remember judgments were not useful for distinguishing between old and new sentences.

For confidence ratings applied to Know judgments, there was a linear abstraction effect for Know hits ($N = 26$), $F(2, 50) = 7.51$, $MSE = 0.470$, $p < .01$, and no linear effect for Know false alarms ($N = 20$), at least with the collapsed data at the high end of the idea-unit scale. The graph, however, in Figure 5 suggests a trend, and a second ANOVA with participants with all four data points was conducted ($N = 12$), producing an F value significant at the .055 level for Know false alarms as well, $F(3, 33) = 2.80$, $MSE = 0.396$. Newman–Keuls tests for the hit data showed confidence ratings for ¾ idea units differing those for 1 and 2 idea unit sentences ($p < .01$), with no difference between 1s and 2s. Newman–Keuls tests for false alarms showed confidence ratings for 3 and 4 idea unit sentences differing from 1s and 2s at the .10 level, with no differences between 1s and 2s, and 3s and 4s. Pairwise t tests showed that participants did not discriminate in their confidence ratings between hits and false alarms, except for the 3-idea-unit comparison, $t(26) = 2.30$, $p < .05$.

Overall, the results on confidence ratings for Remember and Know judgments show similar patterns of increase with semantic integration, although the increase with Know judgments appears somewhat less pronounced. To assess whether Remember judgments per se, independent of the linear abstraction effects, increase confidence ratings, unidirectional pairwise t tests at each idea unit level were conducted comparing ratings of Remember versus Know judgments. The t test comparisons were significant at the .001 level for all comparisons, except the 4-idea-unit comparison, which had fewer participants and only approached significance ($p < .07$).

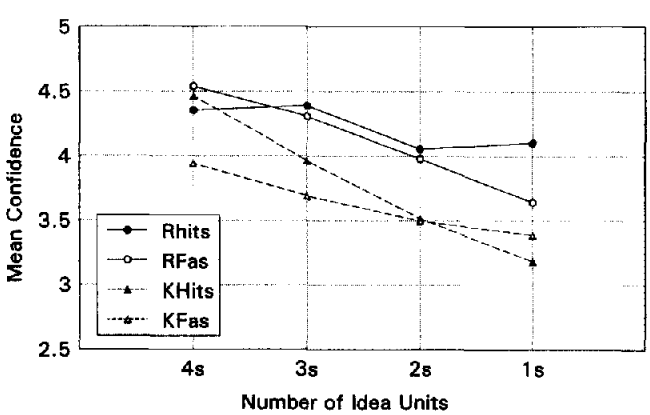

Figure 5. Mean confidence ratings applied to Remember–Know hits and false alarms, Experiment 2. Rhits = Remember hits; Rfas = Remember false alarms; Khits = Know hits; Kfas = Know false alarms.

Discussion

One goal of Experiment 2 was to replicate the general pattern of results seen in Experiment 1, and this was accomplished. Overall recognition, Remember, and Know judgments all showed the same pattern of results as in Experiment 1. This replication indicates two things. First, the results from Experiment 1 are reliable. Second, adding the request for confidence ratings directly subsequent to Remember–Know judgments had no adverse effects, which should encourage the future use of both types of judgments in single experiments.

[3] Limiting the one-way analyses of variance on linear abstraction effects for both Remember and Know judgments to participants with four data points produces equivalent results, but with additional participants dropped from the analyses. The one exception, the analysis on Know false alarms, is presented both ways.

The phenomenological aspects described by Remember–Know judgments share the implications discussed for Experiment 1. Remember judgments, representing memories with episodic content, mirrored the pattern shown by overall recognition, of some discrimination between old and new sentences and linear abstraction effects. Know judgments, however, showed little or no linear abstraction effects and less discrimination between old and new sentences than overall recognition or Remember judgments. Again, the results show (a) Remember judgments primarily driving the pattern for overall recognition, (b) a high number of Remember false memories, and (c) a dissociation between Remember and Know judgments, where Remember judgments appeared more sensitive to the linear effects of semantic integration.

The main goal of Experiment 2 was to directly compare Remember–Know judgments and confidence ratings. We expected confidence to be generally higher for Remember judgments than Know judgments. Furthermore, we expected that as integrative processes increased across sentence complexity levels, confidence would also increase, at least for Remember judgments. Both of these predictions were borne out. In addition, confidence for Know judgments also showed linear abstraction effects suggesting that patterns of confidence are not solely reliant on whether a memory has episodic content. Rather, confidence in one's memory in this paradigm was reliant on the presence of episodic content and the match between the underlying gist representation to recognition items. In the absence of episodic content, a match with the underlying gist representation still provides a basis for linear abstraction effects for confidence ratings. Overall, the results from Experiment 2 suggest that Bransford and Franks's (1971) linear abstraction effect was primarily driven by larger frequencies of Remember judgments, as we previously noted, but that linear abstraction effects for confidence ratings were based on both Remember- and Know-type memories.

If it is true that confidence is based on both the presence of episodic content as well as the degree of match between recognition items and participants' gist representations, then removing or reducing semantic integrative processes should make confidence judgments more reliant on episodic content alone. With reduced integrative pressure there should be less drive to construct an underlying gist representation. This in turn should lessen the preferential reporting of episodic content (Remember judgments) for more complex sentences and reduce linear abstraction effects on recognition memory. With respect to confidence judgments, the absence of a firm underlying gist representation should lessen the linear abstraction effects for both Remember and Know confidence judgments while maintaining higher levels of confidence for memories with episodic content regardless of sentence complexity. This hypothesis was investigated in Experiment 3.

Experiment 3

As indicated above, the goal of Experiment 3 was to test the hypothesis that confidence in Experiment 2 was reliant on the dual processes of episodic content plus the match between recognition items and participants' underlying gist representation. This hypothesis was proposed to account for the fact that both Remember and Know confidence judgments showed linear abstraction effects although only Remember judgments demonstrated such effects in the recognition data. In the previous two experiments, semantic integration processes produced underlying, integrated gist representations. This in turn increased the report of episodic content (Remember judgments) for more complex sentences and appeared to drive the linear abstraction effects found in confidence judgments. In the current experiment, we hoped to reduce levels of semantic integration and demonstrate that confidence would now be solely reliant on the presence of episodic content regardless of level of sentence complexity.

To reduce integrative processes, we used a less semantically biased encoding condition, asking participants to count the number of letters in the last word of each sentence. The goal was not to create an overly powerful perceptual manipulation, but rather to create a perceptual encoding manipulation just strong enough to reduce Remember judgments due to semantic integration effects without seriously affecting overall memory performance. In other words, we wanted the patterns for overall recognition to be similar to those in Experiments 1 and 2. Using this more perceptual encoding manipulation, it was expected that Remember judgments would lose their sensitivity to semantic integration, whereas Know judgments would remain relatively unchanged, or increase slightly, because Know judgments can be more sensitive to perceptual variables. Moreover, it was expected that with the strong effects of semantic integration removed, confidence ratings would be driven solely by whether a memory had episodic content (i.e., those memories identified with Remember judgments). There should be no linear abstraction effects due to the direct match with an underlying gist representation.

Method

Participants. Forty undergraduates served as voluntary participants, receiving experimental credit toward psychology class requirements for their participation.

Materials and procedure. The materials and procedure for Experiment 3 were identical to those from Experiment 2, with the following exceptions: Rather than asking participants to answer elliptical questions after each sentence as a semantic orienting task, participants were instructed to count the number of letters in the last word in each sentence and to write this number on their answer sheet.

Results

The same strategy for analyses was used in Experiment 3 as in the previous experiments. Results for overall recognition are shown in Figure 6. As in Experiments 1 and 2, there were linear abstraction effects for overall recognition hits and false alarms, $F(3, 117) = 7.62$, $MSE = 0.011$, $p < .001$, and $F(3, 117) = 24.16$, $MSE = 0.016$, $p < .001$, respectively. Post hoc Newman–Keuls tests showed only 1s differing from all other idea units for hits ($ps < .05$), and all

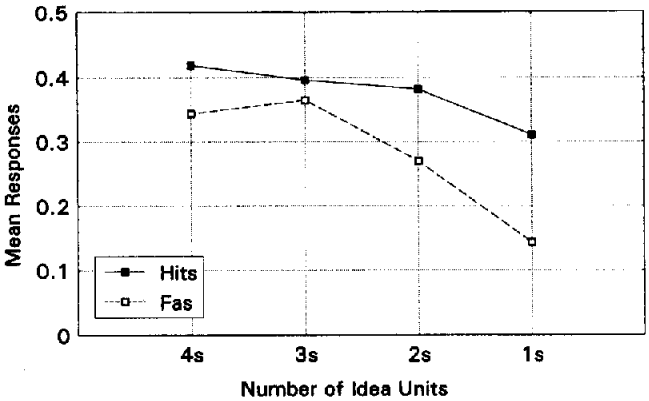

Figure 6. Mean rates of overall recognition hits and false alarms (Fas), Experiment 3.

idea units differing from each other for false alarms ($ps < .05$) except for 3s versus 4s. Pairwise t tests showed participants discriminating between old and new sentences for all idea units ($ps < .05$) except for 3s, $t(39) = 1.55$, $p = .13$. These overall recognition results are generally similar to those from Experiments 1 and 2, though appear to show attenuated linear abstraction effects for both hits and false alarms. To verify this attenuation, we conducted two cross-experiment mixed design ANOVAS on both recognition hit and false alarm rates, examining the change in linear abstraction effects from Experiments 2 to 3. Both analyses produced a significant main effect of linear abstraction across sentence complexity. Of interest to the issue at hand, however, was the presence of a significant interaction between experiment and sentence complexity variables. This interaction was significant for both hits and false alarms, $F(3, 234) = 5.38$, $MSE = 0.011$, $p < .01$, and $F(3, 234) = 3.94$, $MSE = 0.011$, $p < .01$, respectively. Thus, it appears that changing the encoding conditions did have the effect of reducing semantic integration effects. However, because the pattern of results is not overly different from those of Bransford and Franks (1971), it would appear that changing encoding conditions did not completely eliminate semantic processing.

Remember–Know judgments. Results for Remember–Know judgments are shown in Figure 7. For Remember judgments, there were significant linear abstraction effects for both hits and false alarms, $F(1, 117) = 3.44$, $MSE = 0.015$, $p < .05$, and $F(1, 117) = 9.48$, $MSE = 0.012$, $p < .001$, respectively. Nonetheless there appeared to be a flattening of the effect, not unlike that seen for the overall recognition results. Post hoc Newman–Keuls tests showed only 1s differing from 3s and 4s for hits ($ps < .05$), and only 1s differing from 2s, 3s, and 4s for false alarms ($ps < .05$). Pairwise t tests showed participants significantly discriminating between old and new sentences only for 1s and 2s ($ps < .05$). Cross-experiment mixed design ANOVAS verified the impression of attenuation of linear abstraction effects. In addition to the expected main effects of sentence complexity, there were significant Experiment (Experiment 2 vs. 3) × Sentence Complexity effects for both Remember hits and false alarms, $F(3, 234) = 4.87$, $MSE = 0.015$, $p < .01$, and $F(3, 234) = 3.69$, $MSE = 0.012$, $p < .05$, respectively. In these analyses, there were also significant main effects of experiment, with the overall number of Remember hits and Remember false alarms decreasing significantly from Experiments 2 to 3, $F(1, 78) = 6.35$, $MSE = 0.054$, $p < .05$, and $F(1, 78) = 8.43$, $MSE = 0.034$, $p < .01$, respectively. Again, this is encouraging because it shows that the altered encoding condition lessened semantic integration processes, as evident by a reduction in the number of Remember judgments and the attenuation of Remember linear abstraction effects. Indeed, Remember judgment results looked more like Know judgments results.

For Know judgments, there were no linear abstraction effects for hits, though there was a significant effect for false alarms, $F(1, 117) = 6.81$, $MSE = 0.013$, $p < .001$. Post hoc Newman–Keuls tests for false alarms showed only 1s differing from 2s, 3s, and 4s ($ps < .05$). Pairwise t tests showed participants discriminating between old and new sentences only for 1s and 2s ($ps < .05$). These results for Know judgments are similar to those seen in Experiments 1 and 2 and, interestingly, are also similar to the results seen above for Remember judgments. This again suggests that the effects of semantic integration were lessened, resulting in similar patterns for Remember and Know judgments, whereas in Experiments 1 and 2, Remember judgments showed increased sensitivity to integrative processes with stronger linear abstraction effects. We were thus ready to evaluate the question of interest in Experiment 3, specifically, whether reducing the effects of semantic integration would make high confidence more reliant on episodic content per se. If so, there should be a flattening of all linear abstraction effects for confidence, for Remember and Know judgments alike, as well as a continued divergence of confidence for Remember (episodic content) versus Know judgments.

Confidence ratings. Mean confidence ratings for Remember–Know judgments are shown in Figure 8. Once again there were missing data due to some participants not having hits or false alarms across all four levels of sentence complexity for Remember or Know judgments. A similar

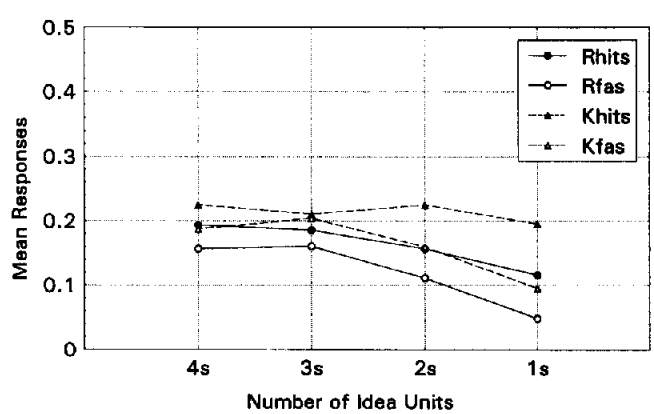

Figure 7. Mean rates of Remember–Know hits and false alarms, Experiment 3. Rhits = Remember hits; Rfas = Remember false alarms; Khits = Know hits; Kfas = Know false alarms.

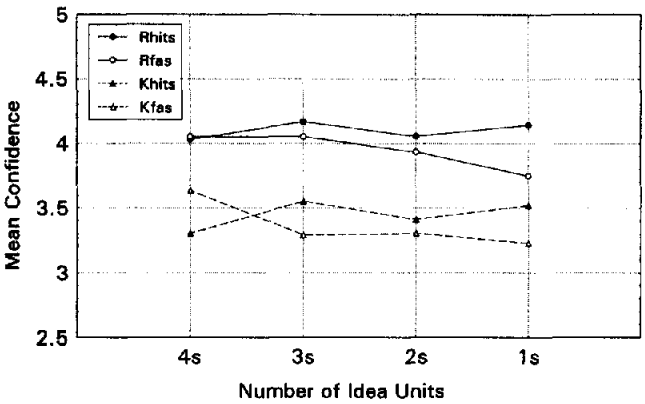

Figure 8. Mean confidence ratings applied to Remember–Know hits and false alarms, Experiment 3. Rhits = Remember hits; Rfas = Remember false alarms; Khits = Know hits; Kfas = Know false alarms.

procedure to that used in Experiment 2 was followed here in collapsing two of the four sentence complexity levels in order to include participants with three of the four data points along the sentence complexity dimension along with those participants with all four points. Not surprisingly, the differential proportion of missing data on the high versus low end of the complexity dimension when comparing Remember versus Know judgments (54% for Remember and 81% for Know on high end) was attenuated when compared with differential rates from Experiment 2 (38% for Remember vs. 81% for Know). A z test on proportions just reached significance levels ($z = 2.03$, $p < .05$, two-tailed). Nonetheless we followed the procedure established in Experiment 2 of collapsing the lower end of the sentence complexity dimension for Remember judgments (1-and 2-idea-unit sentences combined) and the higher end (3- and 4-idea-unit sentences) for Know judgments.[4]

Confidence ratings for Remember judgments showed no linear abstraction effects for hits ($N = 19$) or false alarms ($N = 18$). Nor did pairwise t tests show any discrimination between old and new sentences. Similarly, for Know judgments, there was no linear abstraction effect for hits ($N = 34$) or false alarms ($N = 23$). Pairwise t tests also showed no discrimination between old and new sentences. In sum, the effects of linear abstraction on confidence ratings associated with Remember–Know judgments were eliminated and no differential patterns of confidence ratings across sentence complexity levels for Remember and Know judgments for hits and false alarms emerged. Overall confidence for Remember judgments was, however, greater than overall confidence for Know judgments across all levels of sentence complexity, with all pairwise unidirectional t-test comparisons across the four levels significant, $ps < .001$ for 3s, 2s, and 1s; $p < .05$ for 4s (due to fewer data points).

Discussion

The goals of Experiment 3 were (a) to decrease the effects of semantic integration without overly changing the patterns of results for overall recognition, and (b) to assess whether resulting confidence would be driven primarily by episodic content. Overall recognition showed the same general pattern of results seen in Experiments 1 and 2, with linear abstraction effects for hits and false alarms, and slight discrimination between old and new sentences. Though the general pattern of results was the same, it should be noted that the linear abstraction effects for overall recognition were attenuated in Experiment 3, with the effects primarily driven by differences between most and least complex sentences.

The pattern of results for Remember–Know also showed convergence at higher levels of sentence complexity as compared with Experiments 1 and 2. Remember–Know judgments diverged primarily as sentence complexity increased in the previous experiments, presumably due to Remember judgments being more sensitive to the effects of semantic integration. When the effects of semantic integration were reduced, so was the pattern of increasing Remember judgments with sentence complexity, making Remember and Know judgments behave more similarly. In regard to the phenomenology of memory, the presence of episodic content was more evenly distributed regardless of whether a sentence contained more idea units or not.

The primary goal of Experiment 3, however, was to evaluate the hypothesis that confidence patterns in Experiment 2 were reliant on both the presence of episodic content and the degree of match to the underlying gist representation. If so, then reducing integrative processes should make confidence more, or solely, reliant on the presence of episodic content regardless of level of sentence complexity. The linear abstraction effects for both Remember and Know confidence judgments should be eliminated. Nonetheless, it should continue to be the case that memories lacking distinct phenomenology and supporting memories (Know judgments) are rated with lower confidence. Both of these findings are present in the current data supporting the proposed hypothesis.

General Discussion

The three experiments presented here investigated the phenomenological experience of memories in Bransford and Franks's (1971) semantic integration paradigm. Both Remember–Know judgments and confidence were used to measure phenomenological experience. There were four principle results. First, we replicated Bransford and Franks's (1971) results of poor discrimination and linear abstraction effects using Remember–Know judgments (Experiment 1), and using both Remember–Know judgments and confidence ratings together (Experiment 2). Second, we found that the majority of memories arising from strong integrative pro-

[4] This decision did not affect the results. Findings on linear abstraction effects with the combined data paralleled results when the analyses were conducted with only data from participants who had all four data points along the sentence complexity dimension.

cesses contained episodic content, as measured by Remember judgments. Third, we found large numbers of Remember false alarms, indicating that participants had a conscious awareness of encountering an item with supporting memories that was never presented. Fourth, we found in Experiments 2 and 3 that memory confidence was reliant on both the presence of episodic content as well as the degree of match between recognition items and participants' underlying gist representations. This indicates that although there were larger frequencies of Remember judgments for more complex sentences, these Remember judgments were not wholly responsible for increasing levels of confidence in Bransford and Franks's (1971) study.

Bransford and Franks Revisited

Originally, Bransford and Franks believed that the dual effects of poor discrimination and linear abstraction were due to one underlying memory process. This single process was the construction of an integrated, abstract gist representation of presented sentences. Poor discrimination arose because verbatim sentence structure was never lastingly represented, and linear abstraction effects arose because of the degree of match between underlying gist representations and recognition items (Bransford & Franks, 1971, 1972). The three experiments presented here suggest that this picture is more complex, and indeed more interesting. For one, the majority of memories from semantic integration contained episodic content. Moreover, episodic content accompanied memories that were either true or false, and was not a particularly helpful diagnostic cue in discriminating between true and false memories. Second, though the frequencies of memories with episodic content were responsible for increasing frequencies of memories over sentence complexity (i.e., linear abstraction effects), memories with episodic content were not alone responsible for a rise in confidence over sentence complexity. Confidence was reliant on both the presence of episodic content and the degree of match between underlying gist representations and recognition items.

The reliance on multiple sources for confidence is both encouraging and worrisome. If confidence can be thought of as a consensus, then the more informational sources from which a consensus can derive the better. However, every single base of confidence has the ability to go awry, offsetting the consensus according to its own particular quirks. Moreover, if separate bases of confidence are sometimes influenced by the same variables or task demands, then they have the potential to go awry together, providing a more powerful thrust toward an inaccurate consensus. Indeed, the dual nature of confidence shown in Experiment 2 shows that when two separate bases of confidence are similarly influenced, then overconfidence in inaccurate memories is likely to arise.

Semantic Integration and Episodic Content

Though finding false memories rated with Remember judgments is in general agreement with previous research (e.g., Lane & Zaragoza, 1995; Roediger & McDermott, 1995), the question still remains as to why false memories would have episodic content. We believe the essentially constructive and dynamic nature of memory to be responsible. It is now becoming well established that memory can be largely a constructive process, where retrieved memories are reconstructed entities rather than simple reproductions of past experiences (McClelland, 1995; Moskovitch, 1994; Roediger, 1996; Roediger & McDermott, 1995; Schacter, 1995, 1996; Tulving, 1983). All experiences have a variety of features that may range from gross to particular, such as meaning for the individual, place and time, a particular jacket being red, a particular cup of coffee having a peculiar taste, the sun shining brightly on a particular day, and so on. Although it is probably hard to deny that all these features enter the memory system together, they may not exit the memory system together during recall or recognition (on binding and memory cohesion, see Moskovitch, 1994; Neal Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996).

The first step in Bransford and Franks's (1971) integration paradigm involves the creation of an integrated gist representation, fostered both by variable encoding conditions as well as the very nature of the task, which involves the input of large amounts of material that is associatively very similar and most economically represented as semantically connected. This gist representation is the product of memory construction. As such, it may be that episodic content, which is a complex of certain features of an experience, may become associated, or "bound," improperly. In one case, episodic content could become associated with a gist representation that is not a representation of any particular event.

However, there is another step. At test, participants are required to make decisions concerning whether a particular sentence with a particular sentence structure was presented at study. This is a conflict of task demands, because at study the pressure is to create a gist representation, whereas at test the pressure is for accurate verbatim memory. To state this in a slightly different way, memory was pressured to be constructive at first, and then reproductive. This shift in demands may also contribute to the "drift" or inaccurate binding of episodic content, as the constructed gist representation must be "deconstructed" to some degree to assess whether a particular sentence occurred. Therefore, the opportunity for the particular complex of event features defined by episodic content had at least two opportunities to bind inaccurately with other features, once during encoding and once during recognition.

It may seem odd at first to say that memories of episodes in one's life can be reconstructed. In our experiments, semantic integration created abstractions, summarizations, or the gist of a series of events. Verbatim sentence structure memory would be memory for a specific event. When we have an episodic memory, it is by definition memory of a specific event. However, the majority of memories arising from semantic reconstruction were rated as having episodic content. How could this be true if semantic reconstruction creates abstractions whereas memories for episodes are memories for particular events? From this reasoning, it

would be expected that the majority of memories arising from semantic reconstruction would be rated as not having episodic content, as Know judgments.

The solution to this apparent contradiction lies in the realization again that memory can be a very constructive process. An example will best illustrate this point. Note that for almost every episodic memory, the event can be remembered either from "out of one's own eyes" as it actually occurred, or from some more omniscient point of view, perhaps from somewhere 20 feet away, where the rememberer is actually in the remembered experience in a third person sense (these have been referred as "observer" and "field" memories, respectively (Nigro & Neisser, 1983; Schacter, 1996). Imagine you are sitting in a restaurant eating lunch with a friend. Every so often you look around, perhaps at the waiter, perhaps at the ceiling. You notice the ceiling has an interesting swirly pattern. When asked later to recall the memory of lunch with your friend, there are two possible phenomenological experiences that might arise. First, you might recall the event as you actually experienced it, from out of your own eyes, in which case items in the memory will come one at a time, such that you recollect the waiter, or the ceiling, or your friend. However, you might also recall the event from a more detached point of view, where you envision yourself in the situation. In this case, you recollect the waiter, the ceiling, your friend, and yourself all at once, very unlike what actually occurred. In this latter case, episodic memory was very reconstructive, combining particular events into a representation of a meaningful, context-laden unified event.

From this example, it can be seen that episodic content does not only accompany recollections for specific events, but also accompanies reconstructions of particular events into larger unified wholes, these unified wholes being summarized or abstracted over a number of particular, specific events. Therefore, if inaccurate binding were responsible for episodic content being associated with false memory representations, then the very reconstructive nature of episodic memory itself could create a situation where this inaccurate binding would be more prevalent. This may be why the majority of memories in Experiments 1 and 2 had episodic content, and could be false. As well, it might be the case that, prompted toward semantic reconstruction, memory processes respond with reconstruction in general, not particularly distinguishing between semantic and event reconstruction.

Reality Monitoring and False Memories With Episodic Content

Reality monitoring posits that real memories have on average more contextual and sensory detail and supporting memories than memory for imagined events has (Johnson et al., 1993; Johnson & Raye, 1981). In Experiment 1 we predicted that there would be false memories with episodic content, as reality monitoring theory would predict that, if a person truly believed memory to be accurate, it should have the phenomenological characteristics of actually perceived events. This is however only one possible interpretation of the relationship of reality monitoring to results where false memories are thought to be real memories. A second interpretation is that participants somehow fail to accurately reality monitor (e.g., see Schooler et al., 1986).

This second interpretation however seems like it must be at least partially incorrect in our experiments. This is because failure to successfully reality monitor arises because a person does not effectively attend to reality monitoring cues, such as the presence of sensory and contextual details and supporting memories. Failure to realize a paucity of such cues in a memory may lead one to think the memory is real when it is not. However, as shown by participants actively making Remember judgments on their false memories, it cannot be said that they failed to notice such cues. Making a Remember judgment requires noting such cues. Rather, it would almost appear that participants were very good at monitoring memory cues that should represent real memories; however, these cues were somehow connected to false memories.

A third interpretation is that if memories for imagined memories lack certain phenomenological cues, then the false memories in our experiments should also lack these cues. This is a more problematic interpretation, because false memories in our experiments did not lack those cues, and this finding could be interpreted to be in direct contrast to reality monitoring theory. We suggest however that this interpretation of reality monitoring theory will only be applicable to memories whose features are correctly bound. Indeed, when a false memory comes to be bound with features of real memories, then good reality monitoring on a person's part would mean saying the memory was true because it had the characteristics of a real memory. In sum, reality monitoring theory does accurately encompass our results, though with the additional caveat that paying attention to reality monitoring cues when features of real memories are inaccurately bound to false memories will make one think a false memory to be true.

Conclusion

The present set of three experiments provides an interesting picture of the phenomenology of false memory. All the experiments replicated the Bransford and Franks (1971) effect of poor discrimination and linear abstraction for overall recognition. All the experiments showed the presence of high levels of Remember false alarms, that is, memories with episodic content for events that never happened. As well, Experiments 2 and 3 explored the relationship between episodic content and confidence, finding that confidence was not necessarily solely reliant on the presence of episodic content. These results illustrate the complexities and fruitfulness of phenomenological study of both true and false memory. Apparently, both episodic content and high levels of confidence that a memory is true are not foolproof phenomenological cues as to the accuracy of one's memory.

Our explanations centered on the possible inaccurate binding of episodic content to false memories. We believe this to be a fruitful approach, particularly where there is

great pressure for memory reconstruction. We also suggested that according to this model where the phenomenological features of real memories somehow become inaccurately bound, faithful reality monitoring would actually make one more likely to think a false memory to be real. This suggestion requires further study, as does the relationship between memory binding and reality monitoring in general. If the relationship is accurate, then effective reality monitoring should create more false alarms under conditions where reconstructive demands are high and inaccurate binding is more likely to occur. Under more reproductive memory conditions, effective reality monitoring should prove more diagnostic as to the veridicality of a memory.

References

Angell, J. R. (1909). *Psychology* (4th ed., rev.). New York: Henry Holt.

Baars, B. J. (1988). *A cognitive theory of consciousness.* New York: Cambridge University Press.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2,* 331–350.

Bransford, J. D., & Franks, J. J. (1972). The abstraction of linguistic ideas: A review. *Cognition, 1,* 213–248.

Brewer, W. F. (1992). Phenomenal experience in laboratory and autobiographical memory tasks. In M. A. Conway, D. C. Rubin, H. Spinnler, & W. A. Wagenaar (Eds.), *Theoretical perspectives on autobiographical memory* (pp. 31–50). Dordrecht, The Netherlands: Kluwer Academic.

Cofer, C. N. (1973). Constructive processes in memory. *American Scientist, 61,* 537–543.

Conway, M. A., Collins, A. F., Gathercole, S. E., & Anderson, S. J. (1996). Recollections of true and false autobiographical memories. *Journal of Experimental Psychology: General, 125,* 69–95.

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

Dennett, D. C. (1991). *Consciousness explained.* Boston: Little, Brown.

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

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18,* 23–30.

Gardiner, J. M., & Java, R. I. (1993). Recognizing and remembering. In A. Collins, S. Gathercole, & P. Morris (Eds.), *Theories of memory* (pp. 168–188). Hillsdale, NJ: Erlbaum.

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18,* 579–583.

Griggs, R. A. (1974). The recall of linguistic ideas. *Journal of Experimental Psychology, 103,* 807–809.

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences, 9,* 1–66.

Johnson, M. K. (1988). Reality monitoring. An experimental phenomenological approach. *Journal of Experimental Psychology: General, 117,* 390–394.

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., & 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 errors. *Memory & Cognition, 23,* 607–610.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. F. (1992). When a lie becomes memory's truth: Memory distortion after exposure to misinformation. *Current Directions in Psychological Science,* 1, 121–123.

McClelland, J. L. (1995). Constructive memory and memory distortions: A parallel-distributed processing approach. In D. L. Schacter (Ed.), *Memory distortion: How minds, brains, and societies reconstruct the past* (pp. 69–90). Cambridge, MA: Harvard University Press.

Moskovitch, M. (1994). Memory and working with memory: Evolution of a component process model and comparisons with other models. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 269–310). Cambridge, MA: MIT Press.

Moskovitch, M., Vriezen, E., & Gottstein, J. (1993). Implicit tests of memory in patients with focal lesions or degenerative brain disorders. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (pp. 133–173). Amsterdam: Elsevier.

Neal Kroll, E. A., Knight, R. T., Metcalfe, J., Wolf, E. S., & Tulving, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language, 35,* 176–196.

Nigro, G., & Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology, 15,* 467–482.

Peterson, R. G., & McIntyre, C. W. (1973). The influence of semantic "relatedness" on linguistic integration and retention. *American Journal of Psychology, 86,* 697–706.

Pope, K. S. (1996). Memory, abuse, and science: Questioning claims about the false memory syndrome debate. *American Psychologist, 51,* 957–974.

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

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 365–377.

Rajaram, S., & Roediger, H. L., III. (1997). Remembering and knowing as states of consciousness during retrieval. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to the question of consciousness* (pp. 213–240). Hillsdale, NJ: Erlbaum.

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.

Reber, A. S. (1991). *Implicit learning and tacit knowledge: As essay on the cognitive unconscious.* New York: Oxford University Press.

Roediger, H. L., III. (1996). Memory illusions. *Journal of Memory and Language, 35,* 76–100.

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

Roediger, H. L., III, Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing, and reconstructing the past. In D. L. Medin (Ed.), *The psychology of learning and motivation:*

Advances in research and theory (Vol. 30, pp. 97–134). San Diego, CA: Academic Press.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501–518.

Schacter, D. L. (1995). Memory distortion: History and current status. In D. L. Schacter (Ed.), *Memory distortion: How minds, brains, and societies reconstruct the past* (pp. 1–43). Cambridge, MA: Harvard University Press.

Schacter, D. L. (1996). *Searching for memory: The brain, the mind, and the past.* New York: Basic Books.

Schooler, J. W. (1994). Seeking the core: The issues and evidence surrounding recovered accounts of sexual trauma. Special issue: The recovered/false memory debate. *Consciousness and cognition: An international journal, 3,* 452–469.

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 171–181.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge, MA: MIT Press.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367–447.

Shimamura, A. P. (1986). Priming effects in amnesia: Evidence for a dissociable memory function. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38*(A), 619–644.

Singer, M. (1973). A replication of Bransford and Franks' (1971) "The abstraction of linguistic ideas." *Bulletin of the Psychonomic Society, 1,* 416–418.

Strack, F., & Förster, J. (1995). Reporting recollective experiences: Direct access to memory systems? *Psychological Science, 6,* 352–358.

Strong, E. K. J. (1913). The effect of time-interval upon recognition memory. *Psychological Review, 20,* 339–372.

Suengas, A. G., & Johnson, M. K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined events. *Journal of Experimental Psychology: General, 117,* 377–389.

Tulving, E. (1983). *Elements of episodic memory.* Oxford, England: Clarendon Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–10.

Velmans, M. (1991). Is human information processing conscious? *Behavioral and Brain Sciences, 14,* 651–726.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size constancy. *Journal of Memory and Language, 34,* 622–643.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 934–945.

Received June 10, 1996
Revision received October 2, 1997
Accepted October 3, 1997 ■