

## A Demonstration and Comparison of Two Types of Inference-Based Memory Errors

Sharon L. Hannigan  
Boston University

Mark Tippens Reinitz  
University of Puget Sound

Participants viewed slides depicting ordinary routines (e.g., going grocery shopping) and later received a recognition test. In Experiment 1, there was higher recognition confidence to high-schema-relevant than to low-schema-relevant items. In Experiment 2, participants viewed slide sequences that sometimes contained a cause (e.g., woman taking orange from bottom of pile) but not an effect scene (oranges on floor), or an effect but not a cause scene. Participants mistook new cause scenes as old when they viewed the effect; false alarms to cause scenes and high-schema-relevant items increased with retention interval. Experiment 3 showed that the backward inference effect was accompanied by false explicit recollection, whereas false alarms to schema-high foils were based on familiarity. This suggests that the 2 types of inferential errors are produced by different underlying mechanisms.

There has been growing interest in the cognitive processes that give rise to memory errors. It has become clear that there are numerous independent types of memory errors that arise from different underlying mechanisms. The purpose of this article is to investigate a specific class of memory errors that we call *inferential errors*, which arise as a result of inferences that people make while they experience an event. These inferences can subsequently be mistaken for memories of the actual event.

Researchers interested in how people remember text have extensively investigated a type of inferential error known as *schematic gap filling*. These errors occur when people compensate for their failure to recover specific details about a past episode by drawing inferences and then mistake these inferences for material they had actually experienced (Bower, Black, & Turner, 1979; Bransford, Barclay, & Franks, 1972; Brewer & Treyens, 1981; Johnson, Bransford, & Solomon, 1973; Loftus & Palmer, 1974). The knowledge structures that have commonly been proposed to underlie such inferences are abstract representations of event-typical knowledge in the form of schemas (Bartlett, 1932) or scripts (Schank & Abelson, 1977). These knowledge structures contain entries that represent events that typically occur in a given type of episode; for instance, a script for going to a restaurant would contain entries for being led to a table, receiving menus, ordering food, and so on. One type of evidence for the existence of schemas comes from systematic patterns of distortion in people's memories for specific episodes; in particular, over time, people tend to "remember" an increasing number of events that did not happen during the specific episode, but that are highly typical of that type of episode. As a simple example, participants who read

an account of dining at a restaurant in which stereotypical information such as ordering food had been omitted have a tendency to later incorrectly claim that the information was present in the text (Bower et al., 1979). A related finding was reported by Friedman (1979), who showed that participants who viewed a photographic scene of a specific context (e.g., a child's playroom) later incorrectly remembered that highly context-typical objects (e.g., a teddy bear) had been present in the scene. Although this finding shows memory distortions for individual scenes, rather than for event sequences, it suggests that schematic gap-filling errors occur in memory for both narrative text and visually presented scenes.

Although schematic gap filling has extensively been studied, other types of inference-based memory errors are possible, which have not been as heavily studied. Many authors have pointed out that people have a natural tendency to make inferences regarding the underlying causes of things in the world; for instance, about the reasons for people's behaviors, for their own outcomes, and for unexpected patterns in the environment (e.g., Jones & Davis, 1965; Waldmann, Holyoak, & Fratianne, 1995). If people commonly make causal inferences as they experience events, then these inferences may subsequently be incorporated into their memories for those events. Later, the results of inferential processes might be misremembered as previously experienced components of an external event.

Both schematic gap-filling errors and causal-inference-based errors would be expected given the assumption that a fundamental goal of cognition is to make coherent sense of the world. This assumption was championed by Jenkins and his colleagues (see, e.g., Jenkins & Tuten, 1998), and motivated several pioneering experiments regarding memory for visually presented events. For instance, Jenkins, Wald, and Pittenger (1978) showed participants slide sequences that told a story—for instance, participants saw a series of pictures of a woman making tea. In a subsequent recognition test, participants received some old slides, some "belonging" slides, which were not shown, but that fit the story, and some control slides, that violated the story in some way (e.g., the woman

---

Sharon L. Hannigan, School of Medicine, Boston University; Mark Tippens Reinitz, Department of Psychology, University of Puget Sound.

Correspondence concerning this article should be addressed to Mark Tippens Reinitz, Department of Psychology, University of Puget Sound, 1500 North Warner, Tacoma, Washington 98416. Electronic mail may be sent to mreinitz@ups.edu.

was wearing glasses that never appeared in the original slide sequence). The false-alarm rate was much higher for belonging slides than for control slides. In a second study condition slides were presented in such a way that they told a less coherent story; in this case the false-alarm rate to belonging slides was much lower. Similar findings were obtained by Shaw, Wilson, and Wellman (1986), who showed some participants a set of simple slides that constituted a "generator set" from which it was possible to completely infer the set of rules that gave rise to the individual stimuli. Other participants studied slides that constituted a "non-generator set" that did not provide sufficient information for participants to acquire the underlying set of rules. In a subsequent recognition test, participants who studied the generator set had a high false-alarm rate to stimuli that were consistent with the generating rules, but that had not been studied. The other participant group was unlikely to make false alarms to these items. All of this work is consistent with the notion that a fundamental goal of memory is to encode a coherent story of an event as a whole, rather than simply to encode individually experienced items (the classic work of Bransford & Franks, 1971, is also consistent with this view). Both highly schema relevant and causally implied events would contribute to event coherence; given this view, both types of errors would therefore be expected to occur.

Although the two types of inferential errors may serve similar functions (of maximizing coherence across the individual stimuli), it is unclear whether they are mediated by the same underlying mechanisms. Schemas representing typical episodes may not be useful for making causal inferences. For instance, if a person in a supermarket sees that the floor in the produce section is littered with oranges, she may naturally infer that someone pulled an orange from an unstable area on the stack. Note, however, that such an event is atypical for a supermarket episode, and so would not be likely to be contained in a supermarket script. That is, assuming people commonly make both schematic gap-filling errors and causal-inference-based errors on memory tests, different underlying knowledge structures may give rise to the two sorts of errors.

There were three major purposes in the experiments reported here. The first was to extend the basic finding of schematic gap-filling errors to stimuli that pictorially depict event sequences. Participants viewed a series of slides that depicted typical episodes such as going to a restaurant. They later received a recognition test for the pictures they had seen that included foils that were highly typical of the episode type; we predicted that participants would have a strong tendency to incorrectly endorse such foils.

The second purpose was to demonstrate the existence of a new type of inference-based memory error, called *causal-inference errors*. Participants saw slides of action sequences that contained "effect slides"; for instance, oranges on the supermarket floor or a person cleaning wine from a restaurant table. They later received a recognition test including "cause foils"; these were slides that had not previously been presented and that represented the most likely cause for the effect slides (a person taking an orange from the bottom of the stack or a person knocking over a glass of wine). We expected a higher false-alarm rate for these foils when participants viewed the relevant effect slide during the study phase than when they did not view the relevant effect slide.

The final purpose was to compare the two error types with the goal of determining whether they arose from a single shared

underlying mechanism. Specifically, in Experiment 3 we compared the phenomenal experiences participants had when they made the two types of errors.

## Overview of Experiments

In all of our experiments, participants viewed four different action sequences, or scripts, each of which consisted of photographs of scenes presented in logical sequence. The script headers, or themes, were (a) eating at a restaurant, (b) getting up in the morning, (c) going grocery shopping, and (d) attending a lecture. Immediately following the retention interval, participants received a recognition test that included two kinds of old scenes that were identical to previously studied stimuli and six kinds of distracter scenes that had not been viewed during the study phase. Of the old scenes, half were highly typical of their script (i.e., "old-high") and half were atypical (i.e., "old-low"). The distracter scenes included two types of causal inferential stimuli, controls for each type of causal inferential stimulus, and two types of schema-based stimuli.

*Causal inferential stimuli* were based on "causal sets," each consisting of one cause and one effect stimulus. For example, one of the causal sets for the lecture script was cause: student tipping back in desk; effect: student topples over onto classroom floor. Forward inference test stimuli were effect scenes (i.e., student topples over onto floor) that had not been shown during the study phase and that were contingently related to cause scenes (i.e., student tipping back in desk) that had been viewed during the study phase. Backward inference test stimuli were cause scenes that had not been previously studied, but that were contingently related to effect scenes that had been studied. Forward inference controls involved the presentation at test of effect stimuli without having shown their contingent causes at study, and backward inference controls involved the presentation at test of cause stimuli without having shown their contingent effects at study. Because the only difference between the causal inferential test conditions and their respective controls was the presence or absence at study of causal staging scenes, an increase in false-positive responses to inference stimuli relative to control stimuli can be attributed to causal inferential processing.

Schema-based stimuli were new scenes that were consistent with the scripts shown at study. Schema-high stimuli were highly typical of the scripts they were consistent with, whereas schema-low stimuli were consistent with, but not typical of, the scripts they were consistent with. Because both types of stimuli were about equally visually similar to previously viewed scenes, a finding of more false alarms to schema-high than to schema-low foils would indicate that participants made schematic gap-filling errors. In Experiment 2 we tested for both schema-based and causal-inference errors across retention intervals of 15 min, 24 hr, and 48 hr to demonstrate that recognition errors can result from causal inference and to determine whether the two error types were similarly influenced by retention interval. Finally, in Experiment 3 we investigated the phenomenal experience that accompanied false-positive recognition responses to schema-based stimuli and causal inferential stimuli.

## Experiment 1

The purpose of Experiment 1 was to replicate with scenes the results Bower et al. (1979) obtained using narrative text as stimuli,

and in so doing set the stage for experiments designed to compare schema-based errors and errors based in causal inference. Although there have been many demonstrations of schematic gap-filling errors using narrative text, to our knowledge there has been only one demonstration of similar effects for pictorial stimuli (Friedman, 1979). Moreover, that demonstration was restricted to memory for items present in a single scene, rather than for entire scenes presented as parts of event sequences. In the work of Jenkins et al. (1978) the extent to which belonging scenes were schema relevant was not systematically varied, so it is unclear whether those results provide a demonstration of schematic gap filling per se. In Experiment 1 we tested whether schematic gap-filling errors demonstrated using narrative text also apply to pictures of scenes.

Consistent with Bower et al. (1979), it was expected that participants would be more confident they viewed new schema-high foils than they did new schema-low foils. In addition, we predicted the hit rate to old schema-high scenes would be higher than that for old schema-low scenes because the tendency to remember typical scenes as old should influence responding to both old and new scenes.

## Method

**Participants.** One hundred forty-four Boston University undergraduates participated for credit in their introductory psychology classes. They were tested in 48 groups, each consisting of 3 participants. All participants reported having normal or corrected-to-normal vision.

**Stimulus and apparatus.** The study and test stimuli were 48 color photographic slides of naturalistic scenes taken by Sharon L. Hannigan. Each scene belonged to one of four different action sequences, or scripts, with each script consisting of 12 scenes arranged in logical sequence. The scripts were (a) eating at a restaurant, (b) going grocery shopping, (c) getting up in the morning, and (d) attending a lecture. All of the scenes were presented by using Kodak slide projectors equipped with Gerbrands tachistoscopic shutters. The scenes subtended about 3° of visual angle vertically and about 5° horizontally. The projectors and shutters were controlled by an IBM-AT compatible computer and timing was controlled by a clock card in the computer.

**Design and procedure.** Participants were seated before a white screen in a darkened room. During the study phase of the experiment, they were shown only two of the four possible scripts. Eight scenes (four schema typical and four schema consistent) were presented from each script. The scenes comprising the scripts were presented in sequence for 2 s each and were separated by 3.5-s blank intervals. Four filler scenes that were not related to either of the scripts were drawn from the two unused scripts and served to control for primacy and recency effects, with one appearing at the beginning and one at the end of each script. The two scripts were shown in succession and were separated by a 12-s blank interval. Viewing of the four different scripts was completely counterbalanced so that across subjects all possible script orders were used equally often. In addition, across subjects every schema-high and schema-relevant slide was used equally often as a target (i.e., was presented during the study phase) and as a distracter (was presented only during the recognition test).

The study phase was followed by a 20-min retention interval during which a filler multiple-choice vocabulary test was administered.

After this filler task, participants received an "old"/"new" recognition test that entailed viewing a series of 16 scenes and deciding whether they had seen them during the study phase. Recognition performance was measured by using a 5-point confidence rating scale: A score of 1 indicated that the participant was sure he or she did not see the scene earlier, and a score of 5 signified that the participant was sure he or she did see the scene. Half of the 16 test scenes were old. Two old-high and two old-low stimuli

were drawn from each of the two scripts that had been viewed during the study phase. The new test scenes included four new-high and four new-low stimuli. The old-high and new-high test items were stereotypical of the scripts to which they were relevant (i.e., schema typical), whereas the old-low and new-low stimuli were consistent with, but not typical of, the scripts to which they were relevant. Test slides were assigned to high- and low-typicality conditions on the basis of norms established by Bower et al. (1979). Table 1 shows the assignment of the specific slides to high-typical and low-typical conditions. The table also shows the study and test stimuli that were presented to a specific subject group (note that across groups, all of the high-typical and low-typical slides were used as both targets (old items) and distracters (new items) in the recognition test). Each test scene was presented for 4 s. They were viewed in random order, with the exception that presentation alternated between scenes belonging to the first and scenes belonging to the second script. A different random test order was used for each group of participants.

## Results and Discussion

The median confidence ratings for each of the test conditions are shown in Figure 1. A Wilcoxon test performed on the median confidence ratings indicated there was higher confidence for new schema-high test scenes than for new schema-low test scenes ( $z = 9.20, p < .001$ ). Confidence ratings were also significantly higher for old-high scenes than for old-low scenes ( $z = 4.41, p < .001$ ). These results suggest a bias to respond that highly typical scenes had been previously studied, for both old and new scenes. However, it is possible that better performance for old-high than for old-low scenes simply indicates that the schema-high pictures were generally more memorable than the schema-low pictures. If this were so, then the confidence difference between old-high and new-high scenes should be larger than the confidence difference between old-low and new-low scenes. However, the opposite pattern was true; the difference between old-high and new-high confidence was numerically smaller than the difference between old-low and new-low confidence, and this difference approached significance ( $z = -1.86, p = .053$ ).

The results show that the false memory effects demonstrated by Bower et al. (1979) are not restricted to verbal narratives but are also pronounced when the stimuli are sequences of scenes. In Experiment 1, participants had a tendency to remember pictures that were never presented, but that were stereotypical of the action sequences they viewed. Because those pictures had not been presented, this tendency must have resulted from inferences based in knowledge about the structure of event sequences. The fact that participants mistook their inferences for visual memories extends previous findings regarding schematic gap filling and validates the experimental procedure introduced here as a technique for investigating other potential inferential error types such as causal inference.

## Experiment 2

In Experiment 2, we tested whether people make causal inferences while viewing scene sequences, and then later misremember their inferences as previously presented scenes. The causal stimuli comprised sets of two-scene causal sequences. For example, a cause scene of a two-scene causal sequence embedded in the grocery shopping script might be of a woman plucking an orange from the bottom of an orange stack, with its effect scene being the same woman picking fallen oranges up off the grocery store floor.

Table 1  
Examples of Within-Script Study Sequences and Test Items Used in Experiment 1

Eating at a restaurant	Attending a lecture	Grocery shopping	Morning routine
Study items			
<i>Enter</i>	ENTER ROOM	ENTER STORE	<i>Wake up</i>
<i>Go to table</i>	<i>Look for friends</i>	GET CART	Turn off alarm
LOOK AT MENU	Settle belongings	<i>Look at list</i>	Lie in bed
<i>Discuss menu</i>	<i>Look at other students</i>	<i>Go up and down aisles</i>	<i>Take shower</i>
ORDER MEAL	TAKE OUT NOTEBOOK	Put items in cart	DRESS
Meal arrives	LISTEN TO PROFESSOR	<i>Find fastest line</i>	Fix breakfast
EAT FOOD	Ask questions	WAIT IN LINE	Read paper
Finish meal	CHECK TIME	<i>Put food on belt</i>	BRUSH TEETH
<i>Eat dessert</i>	<i>Close notebook</i>	CASHIER RINGS UP	<i>Comb hair</i>
Bill arrives	<i>Gather belongings</i>	<i>Watch bag boy</i>	Get coat
<i>Leave tip</i>	Stand up	Cart bags out	<i>Get books</i>
LEAVE	Talk	Load bags into car	LEAVE HOUSE
Test items			
BE SEATED	SIT DOWN	PICK OUT ITEMS	Stretch
Put napkins on lap	TAKE NOTES	Get meat	GET UP
PAY BILL	Change position in seat	Talk to other shoppers	Make bed
Drink water	Daydream	PAY CASHIER	EAT BREAKFAST
Go to table <sup>a</sup>	Settle belongings <sup>a</sup>	GET CART <sup>a</sup>	Lie in bed <sup>a</sup>
ORDER MEAL <sup>a</sup>	TAKE OUT NOTEBOOK <sup>a</sup>	Look at list <sup>a</sup>	DRESS <sup>a</sup>
EAT FOOD <sup>a</sup>	CHECK TIME <sup>a</sup>	WAIT IN LINE <sup>a</sup>	Read paper <sup>a</sup>
Bill arrives <sup>a</sup>	Stand up <sup>a</sup>	Cart bags out <sup>a</sup>	BRUSH TEETH <sup>a</sup>

Note. Uppercase items are high-schema relevant, and lowercase items are low-schema relevant. Items in italics were never presented during the test; test items were drawn from each script.  
<sup>a</sup> Items present in study lists.

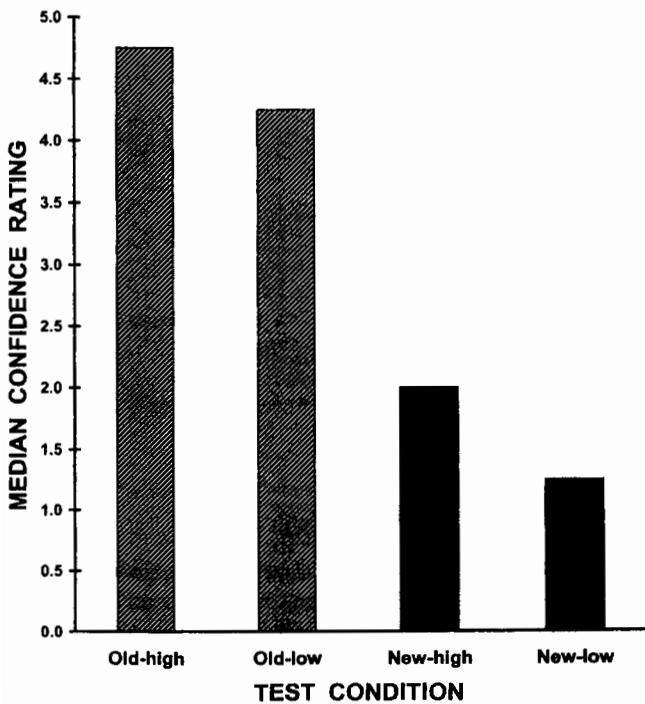


Figure 1. Median confidence ratings in each of the four schema-based test conditions in Experiment 1.

Forward inference errors were said to occur when participants who were shown the cause slide during the study session indicated at test they were confident they had seen the effect slide earlier, when in fact they had not. A backward inference error occurred when only the effect slide of a two-scene causal sequence was shown at study and a participant subsequently endorsed the cause slide at test.

A second objective of this experiment was to replicate the results of Experiment 1 demonstrating higher mean confidence ratings for schema-high test scenes than for schema-low test scenes. We did this to compare the results for the schema-based inferential conditions with the results obtained for the causal inferential test conditions.

The third objective was to see whether the mean confidence ratings for the various inferential conditions would increase significantly as a function of retention interval length. To this end, retention interval was varied across three groups of participants; the specific retention intervals were 15 min, 24 hr, and 48 hr. The manipulation of retention interval length was motivated by the observation that many types of memory errors become more pronounced as the delay between study and test increases, among them the misinformation effect (Belli, Windschitl, McCarthy, & Winfrey, 1992; Ceci & Bruck, 1993), imagination inflation (Goff & Roediger, 1998) and the false fame effect (Jacoby, Kelley, Brown, & Jasechko, 1989). In our own laboratory we recently demonstrated that conjunction errors (i.e., false alarms to new stimuli constructed entirely from previously studied parts) rapidly increased with increasing retention interval (Hannigan & Reinitz, 2000).

## Method

**Participants.** One hundred forty-four Boston University undergraduates participated for credit in their introductory psychology classes. Forty-eight participants were randomly assigned to each of the three retention intervals. They were tested in 48 three-participant groups (16 groups for each retention interval). All participants reported having normal or corrected-to-normal vision.

**Stimulus and apparatus.** The same scenes used in Experiment 1 were again used here, with the exception that study and test stimuli were expanded to include a total of eight two-scene causal sets. There were two different causal sets for each of the four scripts. The two causal sets for the restaurant script were cause: accidentally knocking over a glass of water; effect: mopping up spilt water and cause: finding a hair in the soup; effect: complaining to the waitress. For the lecture script, the causal sets were cause: tilting back in desk; effect: toppling over and cause: called on by professor; effect: gesture, "I don't know the answer". For the grocery shopping script, the causal sets were cause: taking an orange from the bottom of the orange stack; effect: picking oranges up off the floor and cause: grocery bag breaking through; effect: gathering groceries off the sidewalk. Finally, for the morning routine script the sets were cause: fumbling with clock radio alarm; effect: clock radio dangling off night table and cause: catching finger on bowl while scrambling eggs; effect: cleaning up eggs. The same apparatus used in Experiment 1 was also used for this experiment.

**Design and procedure.** For the study phase of the experiment, participants were shown four consecutive scripts in the following order: (a) eating at a restaurant, (b) attending a lecture, (c) going grocery shopping, and (d) getting up in the morning. Each script involved the presentation of 13 scenes arranged in logical sequence, with stimulus exposure duration and interstimulus interval set at 2 s and 3.5 s, respectively. Twelve-second blank intervals separated the scripts. To control for primacy and recency effects, two filler slides preceded the first script shown and two followed the last script shown. The filler slides (pictures of Sharon L. Hannigan's then 2-year-old nephew playing in the woods and sitting in his bicycle seat) were unrelated to the experimental slides.

Embedded in a logically plausible location in each of the scripts was either a cause scene or an effect scene from one of the two causal sets. If a cause scene was viewed, that participant was later tested with its effect scene in the forward inference condition. If an effect scene was presented, that participant was later tested with its cause scene in the backward inference condition. The unused causal set for each script served as forward and backward inference control stimuli such that a cause slide was shown at test without having shown its effect slide at study, and an effect slide was shown at test without having shown its cause slide at study. Across subjects, all of the cause and effect scenes occurred equally often in the experimental and control conditions. Table 2 provides an example of how cause slides were assigned to the various test conditions for a single group of participants.

The study phase was followed by a 15-min, 24-hr, or 48-hr retention interval. Participants who received the 15-min retention interval did word-find puzzles between the study phase and the test; the remaining participants left the laboratory immediately following the study slide presentation and returned at the designated time.

During the test phase, participants were shown 20 slides and were asked to rate their recognition confidence by using the same scale used in Experiment 1. The recognition test included two old test conditions in which there were 4 items and six new test conditions in which there were 16 items. The old test stimuli included two old-high scenes and two old-low scenes. The new test conditions included four schema-high scenes, four schema-low scenes, two forward inference scenes, two backward inference scenes, and two controls for each of the inferential conditions. The controls were included to ensure that high confidence ratings to forward inference and backward inference test stimuli were the result of inferential processing and not due to bias in the forms of stimulus famil-

Table 2  
Examples of the Four Causal Inference Conditions Used in Experiments 2 and 3

Script	Study item	Test item	Condition
Restaurant	Cause 1	Effect 1	Forward inference
	∅ Cause 2	Effect 2	Forward inference control
Lecture	Effect 1	Cause 1	Backward inference
	∅ Effect 2	Cause 2	Backward inference control
Grocery shopping	Cause 2	Effect 2	Forward inference control
	∅ Cause 1	Effect 1	Forward inference control
Morning routine	Effect 2	Cause 1	Backward inference
	∅ Effect 1	Cause 2	Backward inference control

*Note.* Cause 1 = cause slide for Causal Set 1; Effect 1 = effect slide for Causal Set 1; Cause 2 = cause slide for Causal Set 2; Effect 2 = effect slide for Causal Set 2; and ∅ = causal stimulus not shown at study when corresponding member of causal set was shown at test.

ilarity or similarity to scenes previously viewed. Each of the 20 test stimuli was presented for 4 s. Presentation order was random, with the constraint that consecutive test slides were never drawn from the same script. Different random test orders were used for each group, with the constraint that for each 15-min group there was a 24-hr group and a 48-hr group that received the identical study and test items in the identical order. As in Experiment 1, all high- and low-schema-relevant stimuli were used equally often as old and new test items.

## Results and Discussion

As was the case for the Experiment 1 data, we treated the confidence ratings as ordinal, rather than as interval, data. We therefore analyzed differences between medians, using nonparametric tests. Figure 2 shows the median confidence ratings for each of the test conditions for each of the three retention intervals. Pooled over retention interval, Wilcoxon tests showed that we replicated the Experiment 1 findings of higher ratings for old-high than for old-low test items ( $z = 3.45, p < .001$ ), and of higher ratings for new schema-high than for new schema-low stimuli ( $z = 8.42, p < .001$ ).

Again pooled across exposure duration, a comparison of backward inference stimuli with their controls showed that participants had a reliable tendency to make backward causal inferential errors ( $z = 8.13, p < .001$ ). However, there was no evidence that participants made forward causal inferential errors. Surprisingly, confidence ratings were higher to effect scenes in the control condition (i.e., when the cause scene was not studied) than in the forward inference condition (i.e., when the cause had been previously studied) ( $z = 5.05, p < .01$ ). That is, participants had an increased tendency to endorse cause scenes when the associated effect scenes were studied; however, viewing cause scenes during study tended to decrease confidence to effect scenes during the test.

There are several possible reasons for this pattern of effects. We believe it likely that viewing a cause slide during study gave rise to conscious anticipation of the effect slide so that participants tended to notice when the effect slide was absent. Seeing an effect slide should not give rise to anticipation of the cause because slide sequences were always in a forward temporal direction. Moreover, effect slides tended to represent highly noticeable (i.e., unex-

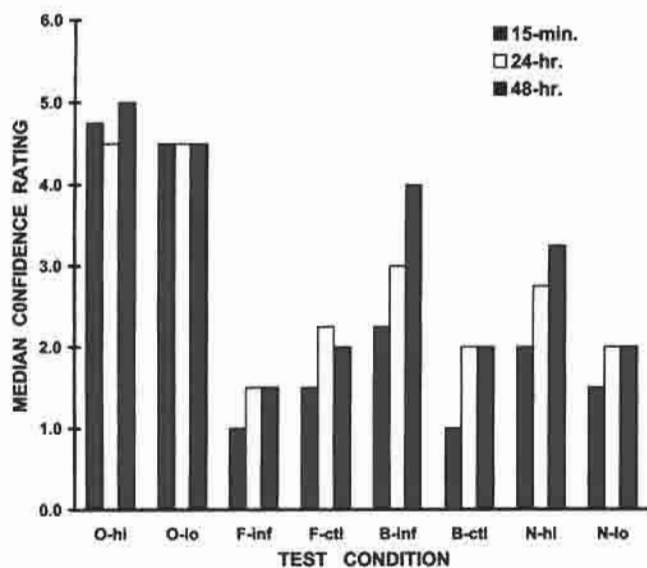


Figure 2. Median confidence ratings in each of the eight test conditions for the three retention intervals in Experiment 2. O-hi and O-lo = old-high and old-low, respectively; F-inf and F-ctl = the forward causal inference and forward causal control conditions, respectively; B-inf and B-ctl = backward causal inference and backward causal control conditions, respectively; N-hi and N-lo = new schema-high and new schema-low test items, respectively.

pected) events. As a result, those slides would be likely to give rise to inferential processing (thereby giving rise to backward inferential errors). By the same token, participants would be more likely to be able to reject such slides at test if they had not been studied because highly noticeable items should be highly memorable. In addition, people may simply be more prone to make inferences regarding cause than regarding effect. It is interesting to note that the entire field of attribution within the domain of social psychology concerns mental processes by which people attribute causes to behaviors and outcomes; there is no corresponding field concerned with automatic inference regarding outcome based on exposure to cause. The lack of existence of a field of study does not imply the lack of existence of an experimental effect. However, the focus by psychologists on attribution may imply a special status for backward inference relative to forward inference.

Old-high confidence increased more with increasing retention interval than did old-low confidence. This indicates that over time there is an increasing bias to respond that anything that is high-schema-relevant is old, with no such bias for low-schema-relevant items. Consistent with this proposal, as retention interval increased, old schema-high items became progressively more difficult to discriminate from new schema-high items, whereas discriminability between old schema-low and new schema-low items remained relatively constant.

Confidence ratings for old-high, new-high, and backward inference conditions progressively climbed as a function of retention interval length. Mann-Whitney tests showed that ratings to old-high items increased significantly from the 24-hr to the 48-hr condition, and ratings for both new-high items and backward inference items increased significantly from 15-min to 24 hr and again from 24 to 48 hr (all  $p$ s < .05). In addition, a measure of the

strength of the backward inference effect was obtained by subtracting the backward inference control score from the backward inference score for each participant. A Kruskal-Wallis test revealed a significant effect of retention interval,  $\chi^2(2, N = 144) = 13.74, p < .01$ , indicating that participants were increasingly likely over time to mistake their causal inferences for actual experiences. Mann-Whitney tests showed that the strength of the backward inference effect increased from 15 min to 24 hr, and again from 24 hr to 48 hr (both  $p$ s < .05).

In contrast, confidence in old-low, new-low, forward inference, and the two inferential controls leveled off following a delay of 24 hr; that is, there were no significant increases in median confidence ratings for these conditions beyond the 24-hr delay (all  $p$ s > .20). This pattern indicates that scores simply do not tend to get higher over time; rather, only those scores specifically related to inferential processing increase over time. The implication is that episodic memory decays over time, thereby necessitating increasing dependence on inference as a way of replacing unavailable episodic information. With this in mind, it is interesting to note that the difference in false alarms to forward inference stimuli and their controls disappeared at the longest retention interval, consistent with the proposal that the difference at shorter durations between these conditions resulted from participants' episodic recollections of their surprise when no effect scene followed the cause scene during study. At the same time, the backward inference effect (i.e., the difference between the backward inference condition and the backward inference control condition) became larger with increasing retention interval, indicating that this effect is quite different from the one that produced a difference between forward inference stimuli and their controls at shorter retention intervals.

It is noteworthy that the patterns of change in the median ratings for backward inference and schema-high conditions were nearly identical across the three sessions, suggesting that a common mechanism may give rise to these inferential error types. The purpose of Experiment 3 was to specifically test this possibility.

### Experiment 3

Experiment 2 showed nearly identical effects of retention interval on mean confidence ratings for backward inference and new-high test conditions. These results provide evidence that both types of errors may be produced by the same underlying mechanism. However, this evidence is weak. Many types of memory errors increase with increasing retention interval (e.g., Ebbinghaus, 1885/1913; Hannigan & Reinitz, 2000; Loftus & Palmer, 1974), and many of these errors have been shown to arise from separate underlying mechanisms.

Moreover, there is additional reason to doubt that the two types of inferential errors we have demonstrated arise from the same underlying source. In particular, schematic gap-filling errors presumably occur because individuals have abstracted the generic sequence of events typical for a given type of episode. In this sense, these errors reflect an individual's within-episode-type knowledge because the event sequence is unique to a particular type of episode. Causal inferences, on the other hand, are not restricted to a particular episode type; that is, leaning back in a chair can cause someone to fall during a meal, during a lecture, while studying, and so on. In this sense such errors are less generic, and more experience specific, than schematic gap-filling errors. The generic, script-specific nature of gap-filling

errors on the one hand, and the less generic, non-script-specific nature of causal inferential errors on the other, raises doubts regarding whether similar underlying knowledge would give rise to the two error types.

We decided to test more rigorously whether gap-filling errors and causal inference errors occur by a single mechanism using a dependent measure that is more sensitive to differences in underlying mechanisms than the measure we previously used. We believed this measure would better illustrate potential differences in the nature of the two error types. Jacoby (1991), Tulving (1985), and others have demonstrated that participants in recognition tests may respond that a test item is old on the basis of a feeling of familiarity in the absence of any specific episodic recollection of the item or from a memory of the specific prior presentation of the item. Tulving devised a method for distinguishing between these two types of recollective experiences by having participants respond that they "know" a stimulus is old when responses are based on familiarity and that they "remember" a stimulus when they specifically remember the prior presentation of the stimulus.

False-positive recognition responses, as well as correct positive responses, may be based in either of these types of recollective experience (see, e.g., Jacoby, Woloshyn, & Kelley, 1989; Reinitz, Morrissey, & Demb, 1994). This idea is consistent with a theory put forth by Tulving (1983) and Tulving, Hayman, and MacDonald (1991), who proposed separate memory systems for generic and episodic knowledge. These authors proposed that recognition responses based in generic knowledge should give rise to "know" responses because such responses should occur on the basis of well-learned (semantic) knowledge that produces familiarity, and that responses based in specific episodic recollection should give rise to "remember" responses. With regard to inferential errors, it seems clear that schematic gap-filling errors should give rise to "know" responses because such errors are based in well-learned generic information. The prediction regarding causal inference errors is less clear. In the previous discussion we pointed out that these errors are more experience specific, and less schema specific, than are gap-filling errors. As a result of the nongeneric nature of these errors, participants may have a tendency to recollect falsely the prior presentation of backward-cause items, thereby making primarily "remember" responses to these items.

To test these predictions we replicated Experiment 2 by using a dependent variable sensitive to the nature of the recollective experiences that accompanied participants' recognition responses. The experiment was run again with a 48-hr-retention interval. However, we used instead the "remember"/"know" procedure devised by Tulving (1985) in the recognition test. By circling *remember* on their response sheets, participants indicated they possessed a clear episodic memory of the test stimulus. A "know" response indicated confidence that a test stimulus was seen in the study phase despite the absence of a specific conscious recollection from the study session, best described as a "feeling of familiarity." If the two error types arise from a common underlying mechanism, then the errors should be associated with the same subjective memorial experience.

## Method

**Participants.** Forty-eight Boston University undergraduates participated for credit in their introductory psychology classes. They were tested

in 16 groups, each consisting of 3 participants. None of the participants had participated in Experiment 1 or Experiment 2. All participants reported having normal or corrected-to-normal vision.

**Stimulus and apparatus.** The study and test stimuli used in Experiment 2 were again used here. They were presented using the same apparatus as before.

**Design and procedure.** The study procedure was identical to that of Experiment 2. Participants left immediately after the study phase and returned to the laboratory 48 hr. later for the recognition test. The test procedure was also identical to that used in Experiment 2, except for the recognition test response options. Participants responded "remember," "know," or "new" to each test stimulus. In this experiment, we utilized the standard instructions used by Gardiner (1988) and Gardiner and Parkin (1990): Participants were instructed to make "remember" responses only if the test stimuli gave rise to some explicit recollection from the study phase; for instance, an image, an association, or some other personal experience. "Know" responses were to be given only if the participant was confident that the test stimulus had been presented during the study phase, despite the fact it evoked no specific conscious recollection from the study phase.

## Results and Discussion

Table 3 shows the frequencies of "remember" responses and "know" responses for each of the test conditions. These data reflect relative response frequencies, rather than confidence. In keeping with all other published studies that have used the "remember"/"know" measures, we performed separate analyses of variance for "remember" and "know" responses. For "remember" responses, the effect of test condition was highly significant,  $F(7, 329) = 59.88$ ,  $MSE = 0.07$ ,  $p < .001$ . For "know" responses, the effect of test condition was also significant,  $F(7, 329) = 4.55$ ,  $MSE = 0.07$ ,  $p < .001$ .<sup>1</sup>

Table 3 indicates that participants tended to "know" that new-high stimuli had been presented and tended to "remember" the prior presentation of backward inference stimuli. Participants made more "know" responses to new-high stimuli than to backward inference stimuli,  $t(47) = 3.74$ ,  $p = .001$ , and they made more "remember" responses to backward inference stimuli than to new-high stimuli,  $t(47) = 2.82$ ,  $p < .01$ .

The main finding of Experiment 3 is the demonstration of a clear difference in the recollective experience accompanying false-positive recognition responses to backward inference and new-high test items. Participants tended to indicate they "remembered" backward inference items and "knew" new-high items. These results suggest that different mechanisms underlie the nearly identical patterns of mean confidence ratings for these conditions observed in Experiment 2. In particular, gap-filling errors are based in familiarity that arises from generic knowledge about

<sup>1</sup> Our analysis included the number of "remember" and "know" responses and assumed the two response types were mutually exclusive. Jacoby, Yonelinas, and Jennings (1997) argued that the response types may be independent, rather than exclusive, and recommended transforming "know" responses to get measures of "pure familiarity" by using the formula:  $F = K/1 - R$ , where  $F$  is familiarity,  $K$  is frequency of "know" responses, and  $R$  is frequency of remember responses. We performed an ancillary analysis using this transformation. The pattern of differences for the transformed measure  $F$  was identical to the pattern for  $K$ . There was no difference in  $R$  and  $F$  scores for backward inference stimuli; however, there was a large difference in  $R$  and  $F$  scores (with  $F > R$ ) for schema-high stimuli.

Table 3  
Mean Frequencies of Remember and "Know" Responses in Experiment 3

Response type	Condition							
	O-hi	O-low	F-inf	F-ctl	B-inf	B-ctl	N-hi	N-low
"Remember"	.81	.70	.07	.18	.49	.08	.29	.08
"Know"	.13	.21	.19	.19	.19	.19	.40	.20

Note. Across subjects all of the causal slides were used equally often in all of the conditions. O-hi and O-low = old-high and old-low, respectively; F-inf, B-inf, F-ctl, and B-ctl = the forward and backward inference and control conditions, respectively; N-hi and N-low = the new high- and new low-schema-relevant conditions, respectively.

event sequences, and causal-inference errors reflect false conscious recollection of the experience of a cause scene that results from the experience of an effect scene.

### General Discussion

#### Summary of Results

There were five major findings. First, the sort of schema-based memory errors for narrative material reported by Bower et al. (1979) and others are easily obtained in a visual recognition test for scenes. Second, we demonstrated the occurrence of a previously unreported memory illusion that results from causal inferential processing. Participants presented with a scene depicting some effect later falsely remembered seeing a slide showing the cause of that effect. Third, in our experiments, causal inferential memory errors occurred commonly in a backward direction, but not in a forward direction. That is, exposure to effect slides led participants to falsely remember having seen the cause, but exposure to cause slides did not induce participants to falsely remember having seen the effect. Fourth, both schema-based and causal-inference-based errors increased with increasing retention interval. Finally, we provide evidence that the underlying memorial processes that give rise to schema-based memory errors are fundamentally different from those that give rise to errors based in causal inference. Specifically, schema-based errors were associated with "know" responses, indicating that they tended to be based in a feeling of familiarity in the absence of explicit recollection (presumably resulting from the activation of generic semantic information). Causally based errors tended to be accompanied by false episodic recollection.

#### Implications of Findings: Gap-Filling Errors and Memory for Typical Events

Gap-filling errors have previously been shown to occur commonly in memory for text, and for objects presented within a scene. The present research shows that such errors also occur when sequences of scenes are used to depict an event sequence and when memory for entire scenes is tested. Our results are consistent with the notion that recognition memory for individual visual scenes in an action sequence is multidimensional. One important component is specific visual information stored when the action sequence is originally experienced. This type of information is assumed to be

primarily responsible for the large overall difference in confidence to old schema-low and new schema-low items. Recognition responses based on stored visual information are manifested as "remember" responses. One would therefore expect new scenes that were visually similar to previously viewed scenes would often receive false-positive "remember" responses.

In addition, the finding that participants make schema-based errors on a recognition test for scenes implies that they sometimes make recognition responses to scenes on the basis of abstract semantic information, rather than on highly specific perceptual information. Because there is no reason to think that schema-low distractors are less visually similar to target items than schema-high distractors are, it follows that semantic consistency, rather than visual similarity, must be responsible for the overall response difference between these types of stimuli. The research provides additional support for the well-known assertion that entries in stored representations of typical action sequences (i.e., scripts) can be mistaken for information experienced during a specific episode at the time of recollection. Moreover, we showed that this is true even when presumably quite memorable pictures serve as stimuli. Interestingly, we demonstrated that these inferential errors are typically experienced by the rememberer as an increase in familiarity, rather than as conscious recollections of previously experienced events.

Finally, we showed that gap-filling errors increase with increasing retention interval. This pattern reflects a decrease in the accessibility of episodic information over time. As episode-specific information becomes more difficult to access, participants become increasingly likely to rely on generic semantic information during the memory test.

#### Implications of Findings: Causal-Inference Errors

We have implicated causal reasoning as a contributing factor to memory errors. Our data suggest that participants engage in such reasoning when they experience events, and that the result of reasoning processes can subsequently be mistaken for previously viewed events. Probably the most interesting finding of this study concerns the quite different recollective experiences that accompanied participants' false-positive responses to schema-based and causal-inference-based test stimuli. We believe this difference indicates that the two error types arise from different sorts of underlying processes. Specifically, schema-based errors result from familiarity that is generated when new high-schema-relevant items activate generic knowledge in memory. We believe that causal-inference-based errors occur when people recall having experienced a specific item during the study phase of the experiment and fail to remember that the experience they are remembering was an inference (an internal event) triggered by a slide in the study sequence, rather than part of the slide sequence itself.

This explanation for causal-inference errors arises directly from the source-monitoring framework advanced by Johnson, Hashtroudi, and Lindsay (1993). Central to their framework is a type of memory error known as *source misattribution*, or confusion about the correct origin of an item or event. One variety of source misattribution is a failure of reality monitoring (Johnson, 1991); that is, a failure to remember whether the source was internal (i.e., self-generated) or external (i.e., phenomenally experienced) to the person. Applied to our findings, episodic memories of inferences

that occurred during the study phase may later have been misattributed to the externally experienced slide sequence. The increase in backward-inference errors with increasing retention interval is consistent with this explanation. According to the source-monitoring framework, memory for specific perceptual attributes experienced during an event is diagnostic on externality. Conversely, the lack of perceptual details may indicate that an item was generated internally. However, memory for specific details is assumed to fade quickly over time. As a result, Johnson et al. (1993) proposed that as the retention interval increases, less perceptual information is required to accept an item as having occurred externally. We believe the reason that our participants were increasingly likely to make backward-inference errors with increasing retention interval is that they were increasingly likely to accept their inferences as externally experienced events as the retention interval increased. Gap-filling errors also increased with the retention interval. However, the mechanism that produced this increase was probably different than the mechanism that gave rise to increased backward-inference errors. In the case of schematic gap filling, loss of episodic information over time tended to leave the effects of familiarity unopposed, resulting in an increase in familiarity-based errors. Unlike gap-filling errors, backward-inference errors tended to be experienced as specific recollections. This implies that backward-inference errors involved the recollection of episodic information, rather than generic semantic information.

An alternative explanation for the backward-inference errors that we observed should be mentioned: It is possible that when participants saw the backward-cause test scenes, they inferred that they must have been presented earlier, and so they responded they were "old" without actually remembering their prior presentation. If this were so, then our findings would constitute a sort of bias effect to respond that scenes consistent with the previous slide sequence were old. However, we think it is unlikely that this is the correct explanation for the backward inference effect. Participants would be expected to have a tendency to make "know," rather than "remember," responses to backward cause items if they were simply inferring that they must have been presented and had no corresponding recollection of their prior occurrence. However, backward-inference errors were almost always accompanied by "remember" responses.

Although we found that participants frequently made backward inference-based errors, we failed to obtain any evidence for the existence of forward inference-based errors. There are several possible reasons for this. As mentioned previously, one possibility supported by the data is that presentation of a cause scene produces conscious anticipation of a subsequent effect scene. As a result, people are surprised when the effect scene is not presented, and later remember the surprising event. In addition, it is likely that effect slides tended to be highly memorable because they tended to depict unexpected events (e.g., oranges all over the supermarket floor). When an effect slide was presented during study, this salience may have given rise to causal inference (resulting in backward inferential errors). When an effect slide was presented at test but not at study, it may have been easy to reject on the basis of its memorability.

There are other possible explanations for our failure to find evidence for forward inferential errors. An appealing possibility is that people are more biased to infer cause from effect than they are

to infer effect from cause. In any case, our finding of strong backward inferential errors lends support to claims by Waldmann and colleagues (1997, Waldmann et al., 1995), who proposed the existence of specific inferential processes (called *diagnostic processes*) that function to attribute effects to likely causes.

### Final Comments

There are applied, as well as theoretical, implications of our results. People may sometimes confuse their schema-based and causal inferences for events that actually occurred. For instance, jury members may remember specific items as having been introduced as evidence, when those items had only been causally implied during testimony. Holyoak and Simon (1999) have provided evidence that inferential processing by jurors can influence their memories. Their participants were asked to reach a verdict in a complex legal case consisting of conflicting arguments. The results indicated that as participants made inferences to reconcile conflicting information in those arguments, their memories for their original positions regarding guilt or innocence shifted. Memories of eyewitnesses may also contain inaccuracies that are the result of inference. For instance, if a witness views a person involved in a bar fight, she may infer the cause of the fight and later mistake that inference for an actual experience. Note that our findings are particularly relevant to courtroom situations because we have shown that the likelihood of inferential errors increases substantially as the retention interval increases. Cases typically go to trial many months after the event occurred; as a result, eyewitnesses may be especially prone to making inference-based errors.

The exact nature of causal inferential illusions of memory remains to be determined. It is possible that inferences, like experienced objects, are represented as units in memory (Reinitz, Lammers, & Cochran, 1992; Reinitz et al., 1994) that can migrate across action sequences, so that, for instance, an inference that occurred in one context (e.g., someone was leaning backward on his or her desk in the classroom) might later engender a false memory about events in a different context (e.g., remembering that someone was leaning backward in his or her chair while studying at home). Alternatively, it is conceivable that inferences are strongly bound to a specific episodic context or even modify the underlying episodic representation of a specific event sequence. Research designed to test between these possibilities is underway.

### References

- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.
- Belli, R., Windschitl, P., McCarthy, T., & Winfrey, S. (1992). Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 356-367.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for texts. *Cognitive Psychology*, *11*, 177-220.
- Bransford, J., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*, 193-209.
- Bransford, J., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, *2*, 331-350.
- Brewer, W., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207-230.

- Ceci, S., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, *113*, 403–439.
- Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Byssene, Trans.). New York: Dover. (Original work published 1885)
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*, 316–355.
- Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.
- Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, *18*, 617–623.
- Goff, L. M., & Roediger, H. L., III. (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory & Cognition*, *26*, 20–33.
- Hannigan, S. L., & Reinitz, M. T. (2000). Influences of temporal factors on memory conjunction errors. *Applied Cognitive Psychology*, *14*, 309–321.
- Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*, 3–31.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *3*, 326–338.
- Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, *118*, 115–125.
- Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Hillsdale, NJ: Erlbaum.
- Jenkins, J. J., & Tuten, J. T. (1998). On possible parallels between perceiving and remembering events. In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole* (pp. 291–314). Washington, DC: American Psychological Association.
- Jenkins, J. J., Wald, J., & Pittenger, J. B. (1978). Apprehending pictorial events: An instance of psychological cohesion. In C. W. Savage (Ed.), *Perception and cognition: Issues in the foundations of psychology* (pp. 129–163). Minneapolis: University of Minnesota Press.
- Johnson, M. K. (1991). Reality monitoring: Evidence from confabulation in organic brain disease patients. In G. P. Prigatano & D. L. Schacter (Eds.), *Awareness of deficit after brain injury: Clinical and theoretical issues* (pp. 176–197). New York: Oxford University Press.
- Johnson, M. K., Bransford, J. D., & Solomon, S. (1973). Memory for tacit implication of sentences. *Journal of Experimental Psychology*, *98*, 203–205.
- Johnson, M. K., Hashtroudi, S., & Lindsay, S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 219–266). New York: Academic Press.
- Loftus, E., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 585–589.
- Reinitz, M. T., Lammers, W., & Cochran, B. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, *20*, 1–11.
- Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 161–168.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- Shaw, R. E., Wilson, B. E., & Wellman, H. (1986). Abstract conceptual knowledge: How we know what we know. In V. McCabe & G. J. Balzano (Eds.), *Event cognition: An ecological perspective* (pp. 59–78). Hillsdale, NJ: Erlbaum.
- Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1–12.
- Tulving, E., Hayman, C., & MacDonald, C. (1991). Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 595–617.
- Waldmann, M. R., & Holyoak, K. J. (1997). Determining whether causal order affects cue selection in human contingency learning: Comments on Shanks and Lopez (1996). *Memory & Cognition*, *25*, 125–134.
- Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General*, *124*, 181–206.

Received March 28, 2000

Revision received November 9, 2000

Accepted November 25, 2000 ■