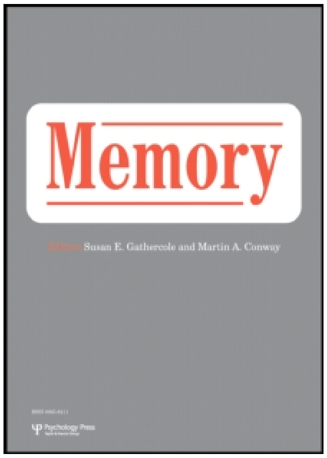

Memory
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683358

Stereotypes influence false memories for imagined events

Heather M. Kleider [a]; Stephen D. Goldinger [b]; Leslie Knuycky [a]
[a] Georgia State University, Atlanta, GA, USA
[b] Arizona State University, Tempe, AZ, USA

First Published on: 12 February 2008
To cite this Article: Kleider, Heather M., Goldinger, Stephen D. and Knuycky, Leslie (2008) 'Stereotypes influence false memories for imagined events', Memory, 16:2, 97 - 114
To link to this article: DOI: 10.1080/09658210801895948
URL: http://dx.doi.org/10.1080/09658210801895948

MEMORY, 2008, 16 (2), 97–114

Psychology Press
Taylor & Francis Group

Stereotypes influence false memories for imagined events

Heather M. Kleider
Georgia State University, Atlanta, GA, USA

Stephen D. Goldinger
Arizona State University, Tempe, AZ, USA

Leslie Knuycky
Georgia State University, Atlanta, GA, USA

Two experiments tested the influences of vivid imagery and person schemata on eyewitness accuracy. Participants watched an event sequence including actors performing stereotype-consistent and inconsistent actions. Additionally, participants either read descriptions of actions (Experiment 1) or vividly imagined actions (Experiment 2). After either 30 minutes or 2 days, recognition memory, source memory, and remember/know judgements were made. After 2 days, false alarms to imagined events increased, relative to the 30-minute test; those false alarms were more often misattributed to stereotype-consistent actors, relative to the same actions in the reading condition. In addition, the accompanying *remember* judgements were higher for false alarms to imagined events, relative to read events, regardless of stereotype consistency. Overall the results suggest that, over time, vivid imagery reinforces schema activation, increasing stereotype-consistent false memories.

Consider the following scenario: A store clerk, working during the holidays, is assisting separate groups of shoppers when she suddenly realises that an expensive purse has vanished from her counter. She contacts a security guard, who asks her to report everything she can recall. While recounting the event, she mentions that a particular girl may have taken the purse, because that purse is popular with pre-teens. However, she cannot be sure, as several people were crowded around the counter, including a teenage boy, an older man, and a couple seeking gifts for their niece. The guard asks her to describe the purse and the girl's appearance, being as detailed as possible. Over the next few days, however, she becomes confident that she *remembers* the girl taking the purse. In this fictional scenario, the clerk uses imagery to help recall events, inadvertently creating feelings of memory for details that were initially based on stereotypical expectations.

In this study we asked whether gender-role stereotypes influence false memories for imagined events. This question is relevant to theoretical issues of eyewitness accuracy and applied issues concerning law enforcement interviewing techniques. Of specific interest is the extent to which vivid imagination promotes the use of stereotypes, leading to source attribution errors. Although there are many bases on which people

Address correspondence to: Heather M. Kleider, Department of Psychology, Georgia State University, Atlanta, GA 30030, USA. E-mail: hkleider@gsu.edu

We thank Peter Killeen for his assistance with the Monte Carlo simulations and Karen Mitchell for her helpful comments on previous versions of this article.

http://www.psypress.com/memory DOI:10.1080/09658210801895948

may attribute memories to sources, the literature suggests that we often rely on memorial attributes or qualitative characteristics, such as perceptual detail, levels of emotional content, or reinstatement of cognitive operations, to link memories to sources (Johnson, Hashtroudi, & Lindsay, 1993). When retrieved details are insufficient for determining the origin of a memory (e.g., Hicks, Marsh, & Ritschel, 2002, Qin, Raye, Johnson, & Mitchell, 2001), people often use top-down knowledge sources, such as schemata, stereotypes, or plausibility (Mather, Johnson, & De Leonardis, 1999; Spaniol & Bayen, 2002). Several studies demonstrate that people use stereotypes when assigning statements to speakers; this occurs more often when the schematic information is provided just prior to test (Bayen, Nakamura, Dupuis, & Yang, 2000; Hicks & Cockman, 2003) or when memory retrieval is difficult (Marsh, Cook, & Hicks, 2006; Mather et. al, 1999). For example Sherman and Bessenoff (1999) found that people used stereotypes in a source memory task (with two possible people, a priest and a skinhead), but only when attention was divided at test. When participants had full attention, stereotype influences disappeared. Stereotypical source errors also increase when memory fades (Spaniol & Bayen, 2002), another situation in which source-specifying information is less available.

More closely related to eyewitness memory, Tuckey and Brewer (2003) found that, after a delay, crime-scene witnesses used their "crime schemata" to interpret ambiguous information, leading to schema-consistent intrusion errors about event details. These false memories were accompanied by feelings of conscious recollection, as indexed by *remember/know* (R/K) judgements. Similarly, Kleider, Pezdek, Goldinger, and Kirk (2008) tested whether schema consistency would influence source memories for depicted actions. They showed participants a slide sequence that included actors performing gender-stereotype-consistent and inconsistent actions, followed (after either 30 minutes or 48 hours) by source memory tests. Given the longer delay, people made more stereotype-consistent source errors, relative to stereotype-inconsistent source errors. In parallel fashion, when people made stereotype-consistent source errors, they more often gave reports of "remembering" (in R/K judgements), relative to trials in which they made stereotype-inconsistent source errors. Together these studies suggest that, even for vivid pictorial events, people sometimes rely on schemata to determine source information and their attributions have the subjective earmarks of true recollection.

In the context of eyewitness memory, it is important to understand the circumstances that promote source errors, for example when determining whether an interview technique promotes accurate retrieval. Several studies have investigated the influence of imagination on event memory; most have focused on eyewitness accuracy, asking whether *guided imagery* in questioning helps or hinders accurate recollection (Henkel, Franklin, & Johnson, 2000). For example, Drivdahl and Zaragoza (2001) investigated whether witnesses, pressed to elaborate upon perceptual characteristics of false event details, were more likely to later falsely remember the fictitious information, relative to a low-imagery comparison group. They found that, after 1 week, people in the high-imagery group were more likely to falsely remember imagined details, and to claim "definite memory" for imagined actions. Drivdahl and Zaragoza suggested that vivid imagery creates perceptual retrieval cues, similar to those of witnessed events, making differentiation between real and imagined events quite difficult. They also expressed concern about police interview techniques in which witnesses (e.g., of a robbery) may be asked to elaborate on details (e.g., about a gun) that were not actually encoded during the event, thus creating inappropriate feelings of certainty during later recall.

The *source-monitoring* framework explains such false memories by suggesting that thoughts, images, and feelings are experienced as memories attributed to particular sources (Lindsay & Johnson, 2001). With vivid imagery, imagined events may acquire the characteristics of truly perceived events, sensory cues, contextual cues (e.g., time and place) and emotional cues (Johnson & Raye, 1981). False memories occur because the created representations closely resemble memories for countless real events. Imagining an event likely makes it feel more available and plausible, thereby increasing its later acceptance as an experienced event (Garry, Manning, Loftus, & Sherman, 1996; Jacoby, Kelley, & Dywan, 1989; see Arbuthnott, Arbuthnott, & Rossiter, 2001, for a review of imagery misattribution effects). Several models have posited why feelings of true recollection sometimes accompany false memories. For example, Lampinen and colleagues suggest *content borrowing*, in which experiential details are borrowed from actual events, providing corroboration for false memories (Lampinen, Meier, Arnal & Leding, 2005; see Henkel & Franklin, 1998, for a reality-monitoring approach). Along similar lines, Brainerd, Reyna, and colleagues suggest that, if strong gist-memory traces closely match false-memory lures, people will experience feelings of true recollection (Brainerd, Wright, Reyna, & Mojardin, 2001; Reyna & Brainerd, 1995).

Although the mechanism is still unclear it appears that, through vivid imagination, people may come to ‘‘remember’’ fictitious details of events. In addition, when recollection is difficult, people may rely on stereotypes when attributing ideas to sources (Mather et al., 1999) or when attributing actors to witnessed actions (Kleider et al., 2008). What has not been addressed in prior literature, and will be addressed in this study, is the extent to which vivid imagery promotes the use of social stereotypes in source recollection. In the present study, we extend the findings of Kleider et al. (2008), now testing whether stereotypical source misattributions, found with witnessed actions, would extend to imagined actions. We used the basic method from Kleider et al. (2008): People were shown a slide sequence wherein two actors performed a series of stereotype-consistent or inconsistent acts. We also asked people either to read descriptions of actions (e.g., *Scott stirs cake batter*), or to both read and *vividly imagine* the action. As in the Kleider et al. study, we tested memory for actions and their associated sources (i.e., which person performed which act), either after 30 minutes or 2 days. We predicted that, with the passage of time, people would experience loss of episodic details, coming to rely more on familiarity to decide whether actions had been seen, increasing false alarms for imagined actions, relative to read actions. Second, we expected that source attributions to all actions (seen, read, or imagined) would show a stereotype-consistent bias. Third, we expected that, when people use vivid imagery for stereotype-consistent actions, they would essentially generate images that borrow from real, or at least schematic, memories. As a result, those generated memory traces would be difficult to distinguish from seen actions, and would encourage the phenomenological experience of remembering.

EXPERIMENT 1: SEEN AND READ ACTIONS

The goals of Experiment 1 were to extend the prior study by Kleider et al. (2008), by adding some actions that were read, and to establish a baseline for comparison to Experiment 2. In Experiment 1, participants watched a slide sequence of two people performing stereotype-consistent and inconsistent actions, intermingled with written actions they were instructed to read. A memory test followed either 30 minutes or 2 days later, manipulated between participants. In both conditions people were instructed to respond ‘‘seen’’ only to actions they *remembered seeing* as pictures. That is, previously read and new actions would all be properly classified as ‘‘new’’. We chose this ‘‘exclusion’’ task because it has natural parallels to eyewitness memory: Imagine that a person has witnessed a crime, and must later recall details. However, she must discriminate things she actually saw from things that were mentioned in the press, or that arose in conversations with friends. Our task was meant to separate truly seen actions from all others.

In Experiment 2 the same photos and written actions were presented, but people were instructed to read, then to vividly imagine, the written statements. We contrasted the reading and reading + imagination conditions across experiments for two reasons: First, we wanted to ensure that participants only read or imagined events, and would not confuse the tasks. Second, we wanted to directly measure the influence of imagery on source memory for actions. Although previous research verifies that people will confuse read and seen actions (e.g., Intraub & Hoffman, 1992, Lane & Zaragoza, 1995), our ultimate goal was to make a clean comparison of simple reading with vivid imagining, using the same actions. Baseline false alarm and source-error rates to read actions were established in Experiment 1.

Method

Participants and design. A total of 75 introductory psychology students at Georgia State University and Arizona State University (35 males, 40 females) participated for course credit. The design was a $2 \times 2 \times 2$ mixed factorial, with Test Delay (30 minutes, $N = 44$; 2 days, $N = 31$) manipulated between participants, Stereotype

Consistency (consistent or inconsistent), and Action Type (read or seen) manipulated within participants. The differences in sample sizes across conditions did not reflect attrition in the Delayed condition; only 2 participants of an original 33 failed to return for the test session.

Materials.. The stimuli included 84 digitised photographs and 15 written action statements. They showed (or described) a woman and a handyman performing various actions, first in the kitchen and then in the backyard of a home. Each depicted action was shown as a set of three sequential pictures, providing a beginning, middle, and end. This ensured that each action was unambiguous. The ''kitchen'' segment included 24 target actions (12 pictorial, 12 verbal). Half of each were stereotype consistent (e.g., the handyman tightening a pipe under the sink); half were inconsistent (e.g., the handyman pouring cake batter into a pan). Examples are shown in Figure 1. There were also six gender-neutral filler actions. The ''backyard'' segment included 13 filler actions that were all stereotype consistent.

The man's attire and actions clearly indicated that he was a handyman, working at the house. The woman's casual attire and actions indicated that she lived in the house. In the opening slides the woman, carrying a small child, answered the front door for the handyman. They then proceeded to the kitchen where she pointed to the sink; he looked at the drain and began working. The sequence continued, with the man performing actions that were consistent for a handyman (e.g., tightening hinges on a cabinet) and the woman performing consistent actions for a homemaker (e.g., folding baby clothes). As the

Figure 1. Examples of stimulus slides. The top and bottom rows show stereotype-consistent and inconsistent action sequences, respectively.

sequence unfolded, both stereotype-consistent and inconsistent actions were performed, or were described in written statements as being performed, by each actor.

To control for action salience, every action was photographed with each person as the actor, with all aspects of the photographs held as constant as possible. Two versions of the sequence were made, counterbalancing actions and actors. This counterbalancing was also maintained for the written action statements. The presentation sequence ended with a stereotype-reinforcing filler scene in the backyard; the homemaker played with two children while the handyman fixed a swing-set. Both actors then returned to the house, the handyman wrote a bill and handed it to the woman. They walked to the front door and each waved goodbye as the man departed.

The selection of stereotype-consistent and inconsistent actions for each actor were validated in a pilot study: Volunteers rated each action on a 1–5 scale, with 1 indicating ‘‘highly expected’’ for the actor. Actions with the most extreme mean ratings, below 2 or above 4 (both consistent, $M = 1.6$, and inconsistent, $M = 4.1$), were selected for inclusion in the picture sequence.

Procedure. Experiment 1 consisted of a presentation phase, a distraction task, then a test phase, all presented on Dell computers with 17-inch CRT monitors. Students participated in groups up to eight people. Each participant sat in a partitioned booth with a computer. Prior to viewing the slide sequence, verbal instructions were given and were repeated in written form on the computer. Participants were instructed that they would see a series of pictures, interspersed with written descriptions, of two people working in a kitchen: Laura (a homemaker) and Scott (a handyman). Their task was to memorise the sequence, in preparation for a later test. They were also instructed to *simply read* the written action statements. The presentation sequence included 84 pictures, each presented for 3 seconds (9 seconds per three-slide action set) and 15 written action statements, each presented for 9 seconds. The total viewing time was 6 minutes, 40 seconds. A series of distractor tasks filled the next 20 minutes, with breaks included to create a 30-minute retention period. The tasks included a novel face recognition test, a number/letter identification task, and a categorisation task.

The test phase followed the distractor phase either immediately or after 2 days. In the test, participants performed an *exclusion* task: 48 action statements (e.g., *stirred cake batter*) were presented one at a time on the computer. The task was to indicate via keypress whether the statement described an action that was *seen* in a picture and, if seen, to indicate which person had performed the action. If the action was either read or new, participants were to respond *new*. Participants were queried regarding their understanding of response types; testing only commenced when all participants indicated understanding the instructions. Finally, for seen responses, participants were asked to make a *remember/know* (R/K) judgement (Gardiner, 1988; Tulving, 1985). Participants were instructed to respond ‘‘remember’’ if they had conscious recollection of seeing the action, and to respond ‘‘know’’ if the action was familiar, but did not evoke a sense of true recollection. When participants classified actions as ‘‘new’’, they did not make R/K judgements. Based on schematic activation, we posited that stereotype-consistent actions would be easily retrieved (or constructed) from memory, based on prior episodes. We assessed whether this extra availability would increase ‘‘remember’’ judgements to stereotype-consistent actions, relative to inconsistent actions. We also anticipated that more ‘‘remember’’ judgements would occur for pictorial actions than read actions, as the memory trace should be more detailed. The action test included 24 statements describing new actions and 24 statements describing old actions (12 previously seen, 12 previously read); half of each were stereotype-consistent.

Results

Analyses of seen, read, and new actions are presented separately. Analyses of seen actions examined two measures: (1) *Hit rates* were defined as the number of truly seen actions correctly called ‘‘seen’’ (regardless of actor), divided by 12. (2) *Source errors* indicated either a single mistake (e.g., the participant correctly recalls that an action was ‘‘seen’’, but recalls the wrong actor), or a compound mistake (e.g., the participant *incorrectly* recalls an action as ‘‘seen’’, when it was actually read, and then also recalls the wrong actor). The calculation of different source error rates is explained below. All effects were considered reliable at an alpha $p < .05$.

Analyses for read and new actions were also conducted on two measures. For previously read actions: (1) *False alarm rates* were defined as the number of read actions incorrectly called ‘‘seen’’, divided by 12. (2) *Source error rates* were defined as the number of false alarms *also* attributed to the wrong actor, divided by the total number of read actions incorrectly called ‘‘seen’’, calculated separately for each participant. For the truly new actions: (1) *False alarm rates* were defined as the number of new actions incorrectly called ‘‘seen’’, divided by 24. (2) *Stereotype-consistent FA rates* were defined as the proportion of false alarms that were attributed to consistent actors. Recognition and source-memory data for seen and read actions are shown in Table 1.

Seen actions. Table 1 shows recognition and source-memory data for actions seen as photos. Note that overall hit rates are given by the sums of ‘‘seen’’ responses, regardless of source attributions. Hit rates, shown in lines 1, 2, 4, and 5 of Table 1, were analysed in a 2×2 repeated-measures ANOVA, testing the within-participants factor Stereotype Consistency and the between-participants factor Test Delay. As one would expect, hit rates were higher in immediate ($M = .86$) than delayed ($M = .74$) testing, $F(1, 73) = 15.4$; $MS_e = .03$; partial η^2 (η_p^2) $= .17$. The main effect of Stereotype Consistency was not significant, nor was its potential interaction with Test Delay.

Of critical importance, we expected stereotype-consistent source errors to increase over time. Another 2×2 ANOVA was performed on source errors (i.e., trials with correct ‘‘seen’’ judgements but incorrect source recall); these data are shown in lines 2 and 4 of Table 1. A main effect of Test Delay was observed, $F(1, 73) = 44.1$; $MS_e = .03$; $\eta_p^2 = .37$, with more source errors in the delayed test ($M = .17$) than in the immediate test ($M = .03$). A main effect of Stereotype Consistency was observed, $F(1, 73) = 8.27$; $MS_e = .03$; $\eta_p^2 = .10$, with more source errors to stereotype-inconsistent actions ($M = .11$) than to consistent actions ($M = .06$). Thus, when people misremembered actors, their errors more often conformed to gender-role stereotypes. An interaction of Stereotype Consistency $\times$ Test Delay was also observed, $F(1, 73) = 5.27$; $MS_e = .02$; $\eta_p^2 = .07$. As shown in Table 1, source error rates did not differ as a function of Stereotype Consistency in the immediate test. All source errors increased over time, but stereotype-consistent errors increased at nearly double the rate of inconsistent errors (as in Kleider et al., 2008), suggesting that, as memory becomes less vivid with the passage of time, people relied more on schemata to make source decisions.

In addition to recognition and source-memory judgements, we collected *remember–know* (R/K) judgements for actions designated as ‘‘seen’’. These judgements provide some insight into the phenomenological experiences that motivated participants’ responses: Clear recollection of actors performing actions was to be indicated by ‘‘remember’’ judgements; a general sense of familiarity with ‘‘know’’ judgements (R/K decisions are related to, but not isomorphic with, confidence ratings; Hirshman & Master, 1997). We did not provide an option for participants to indicate ‘‘guessing’’. Because we were interested in the phenomenological experience of remembering, the analyses focus on remember rates (‘‘remember’’ responses / total responses). Remember rates for seen actions (shown in lines 1–4 of Table 2) were tested by another $2 \times 2 \times 2$ ANOVA, with the factors Stereotype Consistency, Test Delay, and Attribution Type (actions attributed to the consistent or inconsistent actor). This

TABLE 1
Recognition and source attributions to actions seen as photos, Experiment 1

Response/Attribution	*Immediate test*	*Delayed test*
Consistent actor shown		
1 *Seen / Consistent actor	.85 (.03)	.63 (.03)
2 Seen / Inconsistent actor	.02 (.01)	.12 (.03)
3 New	.13 (.02)	.25 (.03)
Inconsistent actor shown		
4 Seen / Consistent actor	.04 (.01)	.22 (.03)
5 *Seen / Inconsistent actor	.80 (.03)	.51 (.04)
6 New	.16 (.03)	.27 (.04)

Asterisks indicate correct responses. Standard errors shown in parentheses. Table line numbers in italics.

TABLE 2
‘‘Remember’’ rates to actions seen as photos, Experiment 1

Response/Attribution	*Immediate test*	*Delayed test*
Consistent actor shown		
1 *Seen / Consistent	.88 (.16)	.59 (.28)
2 Seen / Inconsistent	.02 (.09)	.06 (.10)
Inconsistent Actor Shown		
3 Seen / Consistent	.03 (.08)	.12 (.17)
4 *Seen / Inconsistent	.88 (.18)	.50 (.28)

Asterisks indicate correct responses. Standard errors shown in parentheses.

ANOVA was conducted as a between-participants model, because some participants had empty cells in the design. A main effect of Test Delay was observed, $F(1, 292) = 41.72$; $MS_e = .031$; $\eta^2_p = .13$, with higher ''remember'' rates in the immediate test ($M = .45$), relative to the delayed test ($M = .32$). The main effect of Stereotype Consistency was not reliable, $F(1, 292) = .03$; *ns*, nor was the main effect of Attribution Type, $F(1, 292) = 3.74$, *ns*. However, the interaction of Stereotype Consistency × Test Delay was reliable, $F(1, 292) = 95.97$; $MS_e = .031$; $\eta^2_p = .25$. As shown in Table 2, although all ''remember'' judgements declined over time, this decline was greater for stereotype-inconsistent actions. The interaction of Stereotype Consistency × Attribution Type was reliable, $F(1, 292) = 1030.93$; $MS_e = .031$; $\eta^2_p = .78$, as was the three-way interaction, $F(1, 292) = 95.97$; $MS_e = .031$; $\eta^2_p = .25$. These effects were driven by increased ''remembering'' (over time) for actions attributed to consistent actors.

Read actions. Table 3 shows recognition and source-memory data for actions read as statements. Note that participants were instructed to respond ''new'' to read actions—all ''seen'' decisions were errors. Once these errors were committed, however, participants made source attributions and R/K judgements. For actions that were originally read, we first examined overall false alarms, where people erroneously recalled the actions as ''seen''. People made these errors at surprisingly high, uniform rates. False alarms did not change from immediate ($M = .55$) to delayed ($M = .62$) testing, $F(1, 73) = 1.01$, *ns*. Similarly, false alarms were equivalent across stereotype-consistent ($M = .58$) and inconsistent ($M = .56$) presentation, $F(1, 73) = 1.12$, *ns*. Although neither experimental factor affected overall false alarms, our main focus was on the source attributions that accompanied false alarms: Even when people incorrectly recalled seeing actions that were merely read, they might correctly recall which actors had been stated. When they attributed actions to the wrong person, we classified those as source errors. As shown in Table 3 lines 2 and 4, these ''two-step'' source errors followed the same general pattern as the ''standard'' source errors to seen actions. Fewer source errors occurred in the immediate test ($M = .08$) than the delayed ($M = .28$) test, $F(1, 61) = 34.3$; $MS_e = .04$; $\eta^2_p = .36$. Despite a trend toward more stereotype-consistent errors, the main effect of Stereotype Consistency was not reliable, $F(1, 61) = .92$, *ns*. The trend toward a two-way interaction was also unreliable, $F(1, 61) = .47$, *ns*.

Table 4 shows the ''remember'' response rates for all trials in which people made false alarms to read actions. In this measure people were claiming feelings of true memory for seeing items they merely read. These erroneous ''remember'' judgements were analysed in a 2 × 2 × 2 between-participants ANOVA, with the factors Stereotype Consistency of action, Test Delay, and Attribution Type (action was attributed to the stereotype-consistent or inconsistent actor). No main effects were reliable: ''Remember'' decisions occurred at similar rates in both the immediate and delayed tests (.24 and .26, respectively), and were unaffected by the stereotype consistency of stated actors or whether the action was attributed to the stated actor. Among the potential interactions, only Stereotype Consistency × Attribution Type was reliable, $F(1, 259) = 47.83$; $MS_e = .09$; $\eta^2_p = .16$; participants were more likely to report

TABLE 3
Recognition and source attributions to actions read as statements, Experiment 1

Response / Attribution	*Immediate test*	*Delayed test*
Consistent actor stated		
1 Seen / Consistent	.46 (.04)	.44 (.04)
2 Seen / Inconsistent	.07 (.02)	.22 (.03)
3 *New	.47 (.05)	.34 (.04)
Inconsistent Actor Stated		
4 Seen / Consistent	.09 (.02)	.23 (.03)
5 Seen / Inconsistent	.47 (.05)	.34 (.04)
6 *New	.44 (.06)	.42 (.05)

Asterisks indicate correct responses: Participants were instructed to respond ''new'' to actions presented as text. Standard errors shown in parentheses.

TABLE 4
''Remember'' rates to actions read as statements, Experiment 1

Response / Attribution	*Immediate test*	*Delayed test*
Consistent actor stated		
1 Seen / Consistent	.43 (.40)	.36 (.29)
2 Seen / Inconsistent	.08 (.20)	.14 (.20)
Inconsistent Actor Stated		
3 Seen / Consistent	.09 (.20)	.17 (.29)
4 Seen / Inconsistent	.34 (.36)	.38 (.36)

All results in Table 4 show R/K judgements following recognition errors: Participants were instructed to respond ''new'' to actions presented as text. Standard errors shown in parentheses.

"remembering" an action when they made attributions to the originally stated actors (see lines 1 and 4 of Table 4).

New actions. To provide a baseline for comparison to the read actions, we examined responses to truly new actions. The upper half of Table 5, lines 1 and 2, shows false alarms and source judgements to new actions; the lower half, lines 4 and 5, shows the R/K judgements associated with false alarms. False alarms increased from immediate ($M = .17$) to delayed ($M = .31$) testing, $F(1, 73) = 17.97$, $MS_e = .02$, $\eta_p^2 = .20$. Regardless of Test Delay, when people made false alarms, they were more likely to attribute the actions to consistent actors (.17), relative to inconsistent actors (.06), a main effect of Stereotype Consistency, $F(1, 73) = 52.98$, $MS_e = .01$, $\eta_p^2 = .42$. The Stereotype Consistency × Delay interaction was not significant, $F\ (1, 73) = 0.84$, *ns*. Thus, the results for new actions followed a sensible pattern, with false alarms increasing over time, and mainly being attributed to the "more likely" sources. A similar pattern was observed by Kleider et al. (2008). However, the novel aspect of Experiment 1 was the addition of actions that were read

To determine whether the read actions were effective lures, we compared false alarm rates to read and new actions, collapsing across presented consistency for the read items, as new actions were not associated with any actors. In a 2 × 2 ANOVA with factors Action Type (read, new) and Test Delay, there was a main effect of Action Type, $F(1, 73) = 82.48$, $MS_e = .05$, $\eta_p^2 = .53$, with more false alarms to read actions ($M = .58$) than to new actions ($M = .23$). There was also a main effect of Delay, $F(1, 73) = 7.11$, $MS_e = 4.05$, $\eta_p^2 = .08$, as all false alarms increased over time. The interaction of Action Type × Delay was null, $F(1, 73) = 0.82$, $MS_e = .05$, *ns*. Considering source attributions, a main effect of Stereotype Consistency was again observed, $F(1, 73) = 18.23$, $MS_e = .02$, $\eta_p^2 = .20$, with more consistent than inconsistent attributions ($M = .24$ versus .17, respectively). No interactions (with either Action Type or Test Delay) were reliable.

As before, we examined *remember* rates for new actions (i.e., the proportions of false alarms judged as true "remembering"), as shown in the lower half of Table 5. These data were analysed in a 2 × 2 between-participants ANOVA, with factors Stereotype Consistency (of the erroneous attributions) and Test Delay. We observed a main effect of Stereotype Consistency, $F(1, 136) = 20.65$, $MS_e = .06$, $\eta_p^2 = .13$; people more often "remembered" new actions being performed by stereotype-consistent actors ($M = .34$), relative to inconsistent actors ($M = .15$). There was also a main effect of Delay; people more often "remembered" seeing new actions in the immediate test ($M = .29$), relative to the delayed test ($M = .20$), $F(1, 136) = 4.44$, $MS_e = .06$, $\eta_p^2 = .03$. The interaction between the two factors was not reliable.

Finally, we again compared responses to read and new actions: "Remember" response rates to the read actions (collapsed across Attribution Type) were compared to "remember" rates to new actions. In a 2 × 2 × 2 ANOVA with the factors Action Type (read, new), Attribution Type, and Test Delay, there was no main effect of Action Type, $F(1, 273) = .07$, ns, or Test Delay, $F(1, 273) = 1.15$, *ns*, and the interaction of Action Type × Test Delay was null, $F(1, 273) = 3.19$, *ns*. Considering source attribution type, a main effect of Stereotype Consistency was again observed, $F(1, 273) = 15.91$, $MS_e = .068$, $\eta_p^2 = .06$, with higher rates of "remembering" associated with consistent attributions ($M = .31$), relative to inconsistent attributions ($M = .19$). The interaction of Action Type × Attribution Type was reliable, $F(1, 273) = 4.7$, $MS_e = .068$, $\eta_p^2 = .02$. "Remember" judgements were more likely to follow stereotype consistency among new actions (consistent $M = .28$; inconsistent $M = .15$), relative to read actions (consistent $M = .28$; inconsistent $M = .23$). No other interactions were reliable.

TABLE 5
Recognition, source attributions, and "remember" rates to non-presented (new) actions, Experiment 1

Response / Attribution	*Immediate test*	*Delayed test*
Recognition and source attributions		
1 Seen / Consistent	.13 (.02)	.21 (.02)
2 Seen / Inconsistent	.04 (.01)	.09 (.01)
3 *New	.83 (.02)	.70 (.03)
"Remember" rates:		
4 Seen / Consistent	.40 (.31)	.28 (.24)
5 Seen / Inconsistent	.18 (.25)	.12 (.14)
6 *New	NA	NA

Asterisks indicate correct "new" responses. Standard errors shown in parentheses.

Discussion

The results of Experiment 1 replicated and extended the recent findings of Kleider et al. (2008), showing that, when memories are allowed to decay over a 48-hour delay, people rely more on schematic information to link actors and actions. This pattern was seen in errors to previously seen actions, and to completely new actions. However, the novel aspects of Experiment 1 concerned whether participants would falsely remember having seen actions that were simply read, and to establish rates of stereotype-consistent source errors and judgements of ''remembering''. We found that reading such statements created fairly high (58%) false alarm rates, well above totally new actions (23%). In addition, people were more likely to indicate that they ''remembered'' false alarms to read actions (.50), relative to new actions (.25). These results will provide a baseline for comparison to imagined actions in Experiment 2.

Looking at memory performance over time, the results appear a bit mixed. When actions were originally presented with stereotype-inconsistent actors, source attributions ''shifted'' over time (sampling from different groups of participants) in a stereotype-consistent direction. This ''consistent shift'' led to reliable interactions of Stereotype Consistency × Delay ($p < .0001$ for photographs, $p < .001$ for statements). This consistent shift did not occur for new actions, although a numerical trend was observed. However, to a lesser degree, we also observed the opposite effect, when actions were originally presented with consistent actors. As memorial vividness faded over time, inconsistent attributions tended to increase, and consistent attributions tended to decrease, again leading to reliable interactions ($p < .0001$ for photographs, $p < .05$ for statements). Thus, we observed regression to the mean over time, which complicates interpretation. However, one finding stands out: Looking at all error rates, performance consistently degraded over time, with one clear exception. When people read action statements attributed to inconsistent actors, they committed 56% false alarms in the immediate test, with 47% attributed to the inconsistent actor. In the delayed test, people committed 57% false alarms, but attributions to the ''correct'' inconsistent actor fell to 34%, with the balance shifting towards the consistent actor. This pattern suggests that, although the actions were still familiar after two days, the actor/action associations were less intact, and were subject to schema-based errors.

Beyond demonstrating that schematic expectations affect memory for written items, Experiment 1 sets the stage for Experiment 2. In Experiment 2, people both read and vividly imagined the written statements. By increasing the mental imagery during encoding, we expected people to have considerable difficulty discriminating between viewed and imagined actions (Garry, Frame, & Loftus, 1999; Garry et al., 1996). One might also expect that, by providing more vivid encoding cues, people should gain considerable protection from source errors. Our predictions, however, followed closely from the results of Experiment 1: We predicted that, once memory decisions became harder, people would rely more strongly on schemata. With specific respect to vivid imagery, two potential outcomes seemed likely. Based on the *bizarre imagery* effect (e.g., McDaniel & Einstein, 1986), we might expect stereotype-inconsistent imagery to enhance memory. Alternatively, based on *content borrowing* (Lampinen et al., 2005), we might expect that, even during atypical imagery, people naturally activate and differentiate their imagery from prior, more stereotype-consistent episodes. For example, imagining the statement *''Scott stirs cake batter''* would entail an internal contrast between Scott and Laura, the more expected actor. If so, every trial with unusual imagery might lend incidental familiarity and possibly specific details, to the stereotypical alternative. This would leave participants unsure during source recall, perhaps prompting them to rely on schematic guessing.

EXPERIMENT 2: SEEN AND IMAGINED ACTIONS

Experiment 1 showed that reading statements describing actions can induce people to incorrectly recall seeing those actions, particularly in a delayed test. However, the evidence for stereotype-consistency effects was somewhat equivocal. Using Experiment 1 as a basis for comparison, we modified the reading condition in Experiment 2, in which the same actions were read and then vividly imagined.

Method

Participants and design. A total of 63 introductory psychology students at Georgia State University and Arizona State University (30 males and 33 females) participated for course credit. The design was identical to Experiment 1: It was a $2 \times 2 \times 2$ mixed factorial, with Stereotype Consistency and Action Type (seen, imagined) manipulated within-participants and Test Delay (immediate $N = 27$; two days $N = 36$) manipulated between-participants.

Materials and procedure. The materials and procedure were identical to those in Experiment 1, except that participants were asked to engage vivid imagery for all the written statements. Prior to sequence presentation, the participants were given an imagination exercise to ensure they understood what ''vivid imagining'' entailed. In this exercise, the experimenter read an action statement, similar to those in the sequence, asking each participant to ''take 10 seconds to think about what the actor would be wearing and holding while performing the action, what the object they were using would feel like in his/her hand, and what his/her surroundings would include''. After this imagination exercise, participants described for the experimenter what they had visualised and answered a question about what the imagined actor was holding. This exercise both clarified the task and motivated participants to make a valid effort, as they expected to be questioned during the memory test in a similar fashion. In the recognition test, we followed the ''exclusion'' procedure from Experiment 1: Participants were instructed to respond ''seen'' to actions presented as pictures and ''new'' to everything else.

Results

Analyses of seen, imagined, and new actions are again presented separately, using the same measures and analyses as Experiment 1.

Seen actions. Table 6 shows proportions of ''seen'' responses and source attributions for truly seen actions. All ANOVAs were performed as in Experiment 1. As one would expect, hit rates, shown in lines 1, 2, 4, and 5 of Table 6, were higher in immediate testing ($M = .89$) than delayed ($M = .74$) testing, $F(1, 61) = 16.8$; $MS_e = .04$; $\eta^2_p = .17$. The main effect of Stereotype Consistency was not significant, but we observed an interaction of Stereotype Consistency $\times$ Delay, $F(1, 61) = 4.74$; $MS_e = .02$; $\eta^2_p = .07$. In the immediate test, hit rates were higher to stereotype-consistent actions ($M = .93$) than to inconsistent actions ($M = .84$); these did not differ in the delayed test ($M = .72$ and $M = .75$, respectively).

TABLE 6
Recognition and source attributions to actions seen as photos, Experiment 2

Response / Attribution	*Immediate test*	*Delayed test*
Consistent actor shown		
1 *Seen / Consistent Actor	.89 (.03)	.61 (.04)
2 Seen / Inconsistent Actor	.04 (.01)	.12 (.03)
3 New	.07 (.02)	.28 (.04)
Inconsistent actor shown		
4 Seen / Consistent Actor	.05 (.02)	.27 (.04)
5 *Seen / Inconsistent Actor	.79 (.04)	.48 (.05)
6 New	.16 (.03)	.25 (.03)

Asterisks indicate correct responses. Standard errors shown in parentheses.

As in Experiment 1 we expected stereotype-consistent source errors to increase over time. In the 2×2 ANOVA, a main effect of Test Delay was observed, $F(1, 61) = 17.27$; $MS_e = .07$; $\eta^2_p = .22$, with more source errors in the delayed test ($M = .20$), relative to the immediate test ($M = .05$). A main effect of Stereotype Consistency, $F(1, 61) = 7.31$; $MS_e = .03$; $\eta^2_p = .11$, was found, with more source errors to inconsistent actions ($M = .18$) than to consistent actions ($M = .08$). An interaction of Stereotype Consistency $\times$ Test Delay was also observed, $F(1, 61) = 5.31$; $MS_e = .03$; $\eta^2_p = .08$. As shown in Table 6, source errors did not differ as a function of Stereotype Consistency in the immediate test. All source errors increased over time, but stereotype-consistent errors increased far more than inconsistent errors as shown in lines 2 and 4 of Table 6. Taken together, the results for seen actions closely replicated Experiment 1.

Table 7 shows the ''remember'' response rates for seen actions, shown as a function of delay, stated actors, and participants' source attributions. Among truly seen actions, only a main effect of Delay was observed, $F(1, 244) = 21.94$; $MS_e = .05$; $\eta^2_p = .08$, with higher reports of ''remembering'' in the immediate test ($M = .46$), relative to the delayed test ($M = .34$). Neither the main effect of Stereotype Consistency nor Attribution Type was reliable. Among potential

TABLE 7
"Remember" rates to actions seen as photos, Experiment 2

Response / Attribution	*Immediate test*	*Delayed test*
Consistent actor shown		
1 *Seen / Consistent	.89 (.13)	.60 (.31)
2 Seen / Inconsistent	.04 (.08)	.10 (.18)
Inconsistent actor shown		
3 Seen / Consistent	.05 (.09)	.14 (.17)
4 *Seen / Inconsistent	.88 (.14)	.50 (.34)

Asterisks indicate correct responses. Standard errors shown in parentheses.

interactions, only Stereotype Consistency × Attribution Type was reliable, $F(1, 244) = 9.18$; $MS_e = .06$; $\eta^2_p = .07$, suggesting that participants were more likely to say they "remembered" actions that they attributed to stereotype-consistent actors as shown in lines 1 and 3 of Table 7.

Imagined actions. Table 8 shows proportions of "seen" responses (i.e., false alarms) and source attributions for actions that were imagined. As shown, the imagination manipulation was very effective: The average false alarm rate was 73%, with performance far worse than chance (in Experiment 1, read actions elicited 58% false alarms). In a reversal from Experiment 1, false alarms actually *decreased*, $F(1, 124) = 7.31$; $MS_e = .03$; $\eta^2_p = .11$, from immediate ($M = .77$) to delayed ($M = .67$) testing. This finding is quite interesting, as it suggests that correct rejections are improved by allowing the vivid images to fade over time. Overall false alarms were equivalent across stereotype-consistent ($M = .73$) and inconsistent ($M = .71$) presentation, $F(1, 61) = 0.22$, *ns*. With respect to the source attributions that accompanied false alarms, we found fewer source errors in the immediate test ($M = .10$), relative to the delayed ($M = .29$) test, $F(1, 61) = 20.1$; $MS_e = .05$; $\eta^2_p = .25$. A main effect of Stereotype Consistency, $F(1, 61) = 28.1$; $MS_e = .03$; $\eta^2_p = .32$, reflected higher source misattributions favouring the consistent actor ($M = .27$), relative to the inconsistent actor ($M = .12$). Most important, the Stereotype Consistency × Delay interaction was reliable, $F(1, 61) = 11.4$; $MS_e = .03$; $\eta^2_p = .16$: Source errors increased over time, but predominantly in a stereotype-consistent direction as shown in lines 2 and 4 of Table 8. These results parallel those observed with the actions that were truly seen.

TABLE 8
Recognition and source attributions to read + imagined actions, Experiment 2

Response / Attribution	*Immediate test*	*Delayed test*
Consistent actor stated		
1 Seen / Consistent	.73 (.05)	.49 (.05)
2 Seen / Inconsistent	.07 (.02)	.16 (.03)
3 *New	.20 (.04)	.35 (.05)
Inconsistent actor stated		
4 Seen / Consistent	.13 (.03)	.41 (.05)
5 Seen / Inconsistent	.60 (.05)	.29 (.04)
6 *New	.27 (.05)	.30 (.04)

Asterisks indicate correct responses: Participants were instructed to respond "new" to actions presented as text. Standard errors shown in parentheses.

Table 9 shows the "remember" response rates for imagined actions, shown as a function of the participants' source attributions. People gave "remember" judgements at comparable rates in the immediate ($M = .24$) and delayed ($M = .24$) test conditions, and there was no effect of Stereotype Consistency. However, there was a reliable three-way interaction of Attribution × Stereotype Consistency × Delay, $F(1, 238) = 13.92$; $MS_e = .07$; $\eta^2_p = .06$. This interaction is easily appreciated by examining changes over test delays in Table 9: In the immediate test, "remembering" rates were mainly guided by the actors that were suggested during study (i.e., correct attributions, lines 1 and 4 of Table 9). When consistent actors were imagined, people were 34% more likely to "remember" seeing consistent actors, relative to inconsistent actors. And, when inconsistent actors were imagined, people were 37% more likely to remember inconsistent actors, relative to consistent actors. However, in the delayed test, these patterns changed. When consistent actors had been originally imagined, the phenomenological experience

TABLE 9
"Remember" rates to actions read + imagined statements, Experiment 2

Response / Attribution	*Immediate test*	*Delayed test*
Consistent actor stated		
1 Seen / Consistent	.42 (.33)	.35 (.32)
2 Seen / Inconsistent	.08 (.14)	.14 (.25)
Inconsistent actor stated		
3 Seen / Consistent	.04 (.09)	.24 (.23)
4 Seen / Inconsistent	.41 (.42)	.22 (.24)

All results in Table 9 show R/K judgements following recognition errors: Participants were instructed to respond "new" to actions presented as text. Standard errors shown in parentheses.

of recollection diminished only slightly (21%). Conversely, when inconsistent actors had been imagined the subjective recollective experience was eliminated over time (−2%). This pattern suggests that stereotype consistency affected the phenomenological quality of false memories, but only in the delayed test.

New actions. The upper half of Table 10 shows false alarms and source judgements to new actions; the lower half shows the R/K judgements associated with these false alarms. As shown, false alarms to new actions increased from immediate ($M = .17$) to delayed ($M = .34$) testing, $F(1, 61) = 12.79$, $MS_e = .04$, $\eta_p^2 = .17$. There was also a main effect of Stereotype Consistency, $F(1, 61) = 60.25$, $MS_e = .01$, $\eta_p^2 = .48$, with more false alarms attributed to consistent actors ($M = .20$), relative to inconsistent actions ($M = .06$). The Stereotype Consistency × Delay interaction was also reliable, $F(1, 61) = 10.19$, $MS_e = .01$, $\eta_p^2 = .14$, as the tendency to make stereotype-consistent source errors increased over time.

As before, an important question is how memory for new actions compares to memory for imagined actions—if the imagination manipulation is effective, false alarms to imagined actions should be relatively high. We compared these false alarm rates, collapsing across (presented) stereotype consistency for the imagined actions (because new actions were not presented with actors). There was a large main effect of Action Type, $F(1, 61) = 171.77$, $MS_e = .04$, $\eta_p^2 = .74$, with many more false alarms to imagined actions ($M = .72$) than to new actions ($M = .25$). There was no effect of Delay, $F(1, 61) = 0.96$, *ns*, primarily because Delay and Action Type interacted, $F(1, 61) = 14.10$, $MS_e = .04$, $\eta_p^2 = .19$: Errors increased over time to new actions, but decreased over time to imagined actions. Considering source attributions, a main effect of Stereotype Consistency was observed, $F(1, 61) = 80.40$, $MS_e = .03$, $\eta_p^2 = .57$, with more consistent than inconsistent attributions ($M = .32$ and .13, respectively). There was also an interaction of Action Type and Stereotype Consistency, $F(1, 61) = 4.14$, $MS_e = .04$, $\eta_p^2 = .06$, with more consistent source errors to imagined actions as shown in lines 1 and 2 of Table 10. No reliable interactions with Delay were observed.

TABLE 10
Recognition, source attributions, and "remember" rates to non-presented (new) actions, Experiment 2

Response / Attribution	*Immediate test*	*Delayed test*
Recognition and source attributions		
1 Seen / Consistent	.13 (.02)	.27 (.03)
2 Seen / Inconsistent	.04 (.01)	.07 (.01)
3 *New	.83 (.02)	.66 (.04)
"Remember" rates		
4 Seen / Consistent	.39 (.32)	.32 (.26)
5 Seen / Inconsistent	.13 (.20)	.05 (.08)
6 *New	NA	NA

Asterisks indicate correct "new" responses. Standard errors shown in parentheses.

The lower half of Table 10, lines 4 and 5, shows remember rates for new actions (i.e., the proportions of false alarms judged as true "remembering"). The results mirrored those of Experiment 1, with a main effect of Stereotype Consistency, $F(1, 112) = 36.88$, $MS_e = .05$, $\eta_p^2 = .25$. People more often "remembered" new actions as being performed by stereotype-consistent actors ($M = .35$), relative to inconsistent actors ($M = .09$). There was a marginal effect of Delay, $F(1, 112) = 3.35$, $p = .07$, $MS_e = .05$, $\eta_p^2 = .03$, with more "remember" judgements in the immediate test ($M = .26$) than the delayed test ($M = .18$). The two-way interaction was not reliable.

COMPARISON OF EXPERIMENTS 1 AND 2

In this study, our primary goal was to investigate the influence of imagery on schematic memory. Across experiments, the seen and new actions were constant—hit rate to seen actions, $F(1, 134) = ns$; false alarm rate to new actions, $F(1, 126) = ns$—the critical contrast was carried by the read (Experiment 1) and imagined actions (Experiment 2). Because the same actions were presented to participants in both experiments, they allow direct comparison. We first examined false alarms and source attributions across experiments (combining data from Tables 3 and 8), using a 2 × 2 × 2 between-participants ANOVA. In overall false alarms there was a main effect of Experiment, $F(1, 536) = 12.0$, $MS_e = .053$, $\eta_p^2 = .02$, as reading actions elicited fewer false alarms ($M = .29$), than imagining actions ($M = .36$).

Considering source attributions, Figure 2 shows "two-step" source errors (i.e., false alarms to read or imagined actions, which were *also* attributed to the non-stated actor), as a function of presentation consistency and delay. To normalise the data, each value is shown as a

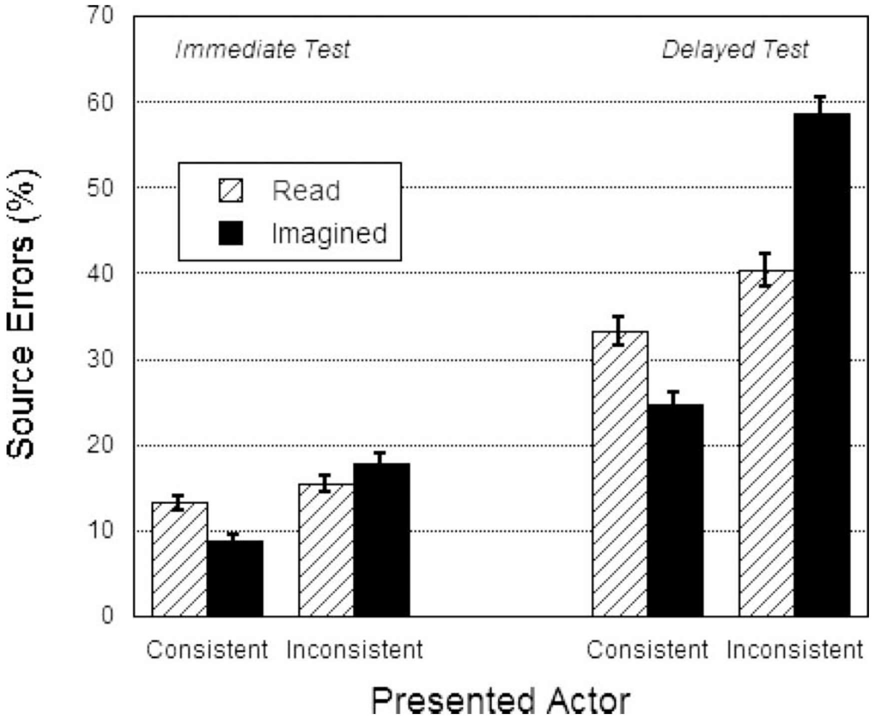

Figure 2. Source errors, shown as percentages of total false alarms, in Experiments 1 and 2. Results are shown for read (striped bars) and imagined (solid bars) actions, in both the immediate and delayed tests. Error bars represent standard errors.

percentage of the total false alarms in its respective condition. These data were tested in a mixed-model, repeated-measures ANOVA. As the figure suggests, a main effect of Delay was observed, $F(1,\ 134) = 51.46$, $MS_e = .03$, $\eta_p^2 = .28$. There was also an effect of Stereotype Consistency, $F(1,\ 134) = 19.09$, $MS_e = .03$, $\eta_p^2 = .13$, with more errors to actions presented with inconsistent actors. Although more source errors occurred with imagined actions, the effect of Action Type was only marginal, $F(1,\ 134) = 3.49$, $MS_e = .03$, $p < .06$, $\eta_p^2 = .03$. Most important, the three-way interaction of Action Type × Stereotype Consistency × Delay was reliable, $F(1,\ 134) = 6.68$, $MS_e = .03$, $\eta_p^2 = .05$. As shown in Figure 2, as the consistency effect grew considerably larger in the delayed test, but primarily for the imagined actions. Apparently, the increased imagery in Experiment 2 did not help people reduce schematic memory errors.

As noted in the discussion of Experiment 1, our results showed stereotype-consistent memory changes over time, but also smaller degrees of memory changes in the opposite direction. Thus, changes in stereotype consistency effects are difficult to separate from simple regression to the mean—as recognition accuracy declines over 48 hours, some portion of the new mistakes are allotted to each category. Moreover, in our design, different participants received the immediate and delayed tests, so all inferences of memory "changes" over time are based on sampling theory, not direct observation. The question naturally arises, then, whether the apparent stereotype-consistent memory changes are due to chance fluctuations.

To address this question we applied a bootstrapping (or *resampling*) procedure (e.g., Chernick, 1999). For both Experiments 1 and 2 we randomly sampled participants from the immediate and delayed testing conditions, joining them together to create "virtual participants" who were represented in both conditions. Once the random pairings were completed, we computed "consistent shift" scores: For each virtual participant, we measured the net changes (from immediate to delayed testing) in both stereotype-consistent and inconsistent source errors. Note that, because we randomly paired different real participants, such "changes" over time could be positive, negative, or nil. We then computed difference scores (consistent minus inconsistent). Again, these difference scores could range from negative to positive values: When positive, they indicated that stereotype-consistent errors increased over time more than stereotype-inconsistent

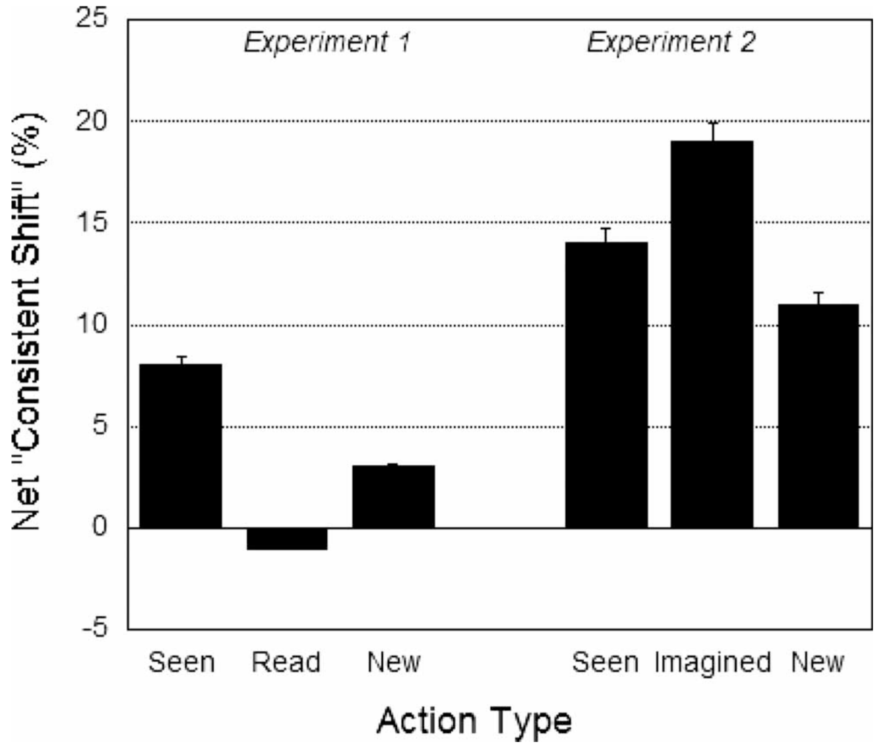

Figure 3. Net "consistent shifts" (shown as percentages) in source errors over time in Experiments 1 and 2. Error bars represent standard errors.

errors. Finally, to generate reliable estimates, we repeated the random pairings and computations 1000 times, in a Monte Carlo procedure, then calculated the average "consistent shift" scores for each condition. The results are shown in Figure 3.

We analysed these data in two ways. First, we conducted one-sample *t*-tests, testing whether the net shifts differed from zero. In four of six cases, there were reliable shifts in the stereotype consistent direction (all $p < .01$); only the *read* and *new* actions in Experiment 1 failed to surpass the significance criterion. Second, we contrasted Experiments 1 and 2 in two repeated-measures ANOVAs. We first compared all three action types across experiments, finding a main effect of Experiment, $F(1,\ 134) = 28.05$, $MS_e = .04$, $\eta_p^2 = .22$. As shown in Figure 3, there were generally larger "consistent shifts" in Experiment 2. We also observed an interaction of Action Type × Experiment, $F(1,\ 134) = 23.12$, $MS_e = .04$, $\eta_p^2 = .20$, reflecting the dramatic change across read and imagined actions. In the second ANOVA we removed the read and imagined actions, leaving only the seen and new actions (which were identical across experiments). Another main effect of Experiment was observed, $F(1,\ 134) = 16.95$, $MS_e = .05$, $\eta_p^2 = .16$, with no interaction. Engaging in vivid imagining not only made overall recognition more difficult, it also fostered greater reliance on stereotypes. Taken together, the contrasts of Experiments 1 and 2 showed that vividly imagining actions led to stereotype-consistent memory errors, similar to those found with seen actions, and at higher rates than reading those same actions.

GENERAL DISCUSSION

The goals of this study were to determine the extent to which schematic activation, vivid imagery, and the passage of time promote source memory errors for witnessed actions. We expected that, given delayed testing, people would rely on the strength of encoded sensory cues to differentiate between actions that were seen and actions that were only read or imagined. As the imagination task likely generates stronger sensory cues, we expected false recognition to be prevalent after the imagery task (Garry et al., 1996, 1999). Moreover, we expected that engaging imagery would activate ingrained stereotypes, perhaps more strongly than reading, leading to strong stereotype consistency effects in source memory. After 2 days, the combination of imagined sensory cues and activated stereotypes should result in schema-consistent source attribution errors. The results supported these predictions: After a 48-hour delay, when imagined events were falsely remembered as being seen, they were most often attributed to the stereotypical actors. This pattern occurred more clearly for imagined actions that for new, read, or truly seen actions. Together these results suggest that schematic processing and vivid imagery both contribute to the creation of false event memories and source errors.

Previous research shows that *reality-monitoring* errors (i.e., discriminating external and internal sources of ideas) increase when people engage in vivid imagery, suggesting that the rich perceptual information associated with externally experienced events (e.g., the sensation of holding a key while unlocking a door) also becomes associated with imagined events (Henkel & Franklin, 1998; Johnson, Foley, & Leach, 1988; Johnson, Raye, Foley, & Foley, 1981; Mammarella & Cornoldi, 2002). Comparatively, internally generated events often contain more information regarding cognitive operations; to reduce reality-monitoring errors, a person must remember either contextual details of encoded episodes or details of his/her own processing of the episodes (Kensinger & Schacter, 2006). We expected that, once event memories had faded, people would struggle to recall such contextual details, instead making recognition and source decisions by relying on overall familiarity and schemata.

As noted earlier, it was arguable whether vivid imagination would lead to source errors in immediate testing, regardless of stereotype consistency, although we expected delayed testing to promote such errors. Against our prediction, relative to reading action statements, imagery allows deeper encoding (Craik & Tulving, 1975) and, in stereotype-inconsistent cases, it allows the memorial benefits of bizarre imagery (Baddeley & Andrade, 2000). Thus, we might have expected people to make false alarms (reporting ‘‘seen’’ for imagined actions), but to keep the imagined actors straight. However, several factors seem to have acted in combination, making imagined actions highly susceptible to schematic processing. First, when given the instructions to vividly imagine some action (e.g., *repairing a drain*), people likely activated prior memories, potentially reinforcing the stereotype-consistent connection between one of the actors and the action. Those activated memories would naturally influence ‘‘construction’’ of the new memory trace. Lyle and Johnson (2006) suggested that false memories can import specific details from actually experienced events; such detail-embellished false memories are subjectively similar to experienced events. Second, in Experiment 2, participants were required to imagine several actions with both consistent and inconsistent actors. These imagination exercises were interspersed with trials requiring memorisation of photographed actions, again with both consistent and inconsistent actors. This situation creates considerable opportunity for confusion. Given an immediate memory test, when the actor–action links were more intact, false alarm rates were very high (72%) to imagined actions, but source errors were relative rare. Given a delayed test, two interesting changes occurred among the imagined actions: Overall correct rejections actually improved, as their vivid images apparently faded. And, for those actions that still elicited false alarms, stereotype effects increased. Although our findings are contrary to some prior literature showing that bizarre stimuli are highly memorable, we suggest that, in the present context, our stereotype-inconsistent actions were not bizarre enough to confer a memorial advantage.

Prior studies on schematic memory errors have focused on the interactions of general knowledge and episodic details in memory traces. For example, Mather and Johnson (2003) found that, relative to young adults, older adults made more stereotypic errors when recalling sources of statements, purportedly made by either Republicans or Democrats. Mather and Johnson suggested that natural declines in executive function will engender greater reliance on schemata. Similar results were reported by Sherman and Bessenoff (1999), who tested people performing a source-memory task under divided attention. Bayen et al. (2000) found profession-biased source errors when people attributed statements to sources (either doctors or lawyers). Bayen et al. attributed their findings to a *guessing hypothesis*, in which people guess based on schemata when they cannot remember information sources. As noted, our experiment was quite challenging. Schematic guessing certainly might have affected our results: For example, schema-consistent errors occurred even for new actions. Thus, people might have combined truly remembered details with schemata to make source judgements. Hicks and Cockman (2003) attempted to isolate these effects by making profession-related information available to people either during encoding and retrieval, retrieval only, or not at all. Their results showed that source decisions were biased by schemata in the retrieval-only condition. They also found that response biases were not simply due to guessing but were more likely when confidence ratings were high. However, Mather et al. (1999) found that high confidence is not necessarily indicative of recollection, as high confidence is at times associated with source memory for new items.

High confidence in source-memory errors have been described in terms of the *source-monitoring framework* (Johnson et al., 1993; Mitchell & Johnson, 2000), in which people make source decisions based on qualitative characteristics bound in memories when events are experienced, such as episodic contextual features. When retrieving information, people sample features of memory traces to determine their origins (internally generated or externally experienced). This process, however, naturally activates schematic frameworks used for the judgements. If a memory trace seems to "fit" with a schema, it may be recalled as an externally experienced episode. In our study, stereotype-consistent errors were more likely for both pictures and imagined events after a delay, a finding consistent with the source-monitoring framework. However, stereotype-consistent errors were not inflated for actions that were read, suggesting that vivid imagining more strongly activated the gender schemata. Falsely recalled imagined events were also more likely to receive "remember" judgements, relative to read events, although stereotype consistency effects were weak. These findings are similar to those reported by Macrae, Schloerscheidt, Bodenhausen, and Milne (2002), who found that, when people made profession-consistent source errors, they were typically accompanied by the subjective experience of "knowing".

Several models have been proposed to account for the recollective experience of false memory. Lampinen and colleagues proposed a *content borrowing* model in which experiential details are borrowed from actual events, providing corroboration for false memories (Lampinen et al., 2005; see also Henkel & Franklin, 1998, for a reality-monitoring approach). This occurs when activated familiar events result in a biased search for episodic details that corroborate the feeling of familiarity. Although we did not find an increase in "remember" responses for imagined stereotype-consistent actions, the content borrowing model and the biased search hypothesis both explain how the stereotype-consistent memory errors occurred for pictures and imagined events, but not for read events.

Brainerd and colleagues' *fuzzy trace theory* (Brainerd et al., 2001; Reyna & Brainerd, 1995) is another approach that may explain the recollective experience of false memories. Stated in terms of our study, how can people falsely "remember" seeing a stereotype-consistent actor performing an action when they actually saw the inconsistent actor, or perhaps saw nothing? According to fuzzy trace theory, memory traces vary on a continuum of specificity, from gist to verbatim. Gist traces represent the general sense and meanings of items; verbatim traces represent more specific item-level information. True memories can be recalled by retrieving either verbatim traces or gist traces that closely match the test item. Conversely, false alarms are produced by retrieving gist representations that closely match test lures. False feelings of recollection occur when the underlying gist trace is strong and is an especially good match for the lure. Sometimes people will confabulate experiential details (e.g., sensory or contextual information) that make false memories seem subjectively real (see Lyle & Johnson, 2006). In the case of imagined events in our study, the sensory and contextual cues were already part of the episodic trace.

According to fuzzy trace theory, participants in the immediate test condition should have mainly retrieved verbatim traces when presented with test statements. As a result, they were fairly accurate in source attributions. After a delay, verbatim traces lose fidelity and/or become less available, leading to greater reliance on gist traces. By their nature, gist traces are likely to reinforce schematic information, including likely actors for actions. Reliance on gist traces would thus increase stereotypical source errors, as we observed.

From a memory-processing perspective, the present results suggest that, when episodic memory is less vivid, people rely on a combination of retrieved sensory/contextual cues, supplemented with schema-consistent details. This strategy is often accurate, as vivid sensory cues are often normally linked to events experienced first-hand. Moreover, schematic (general) information will typically support accurate recollection. Imagine, for example, a person who sees her doctor and hairdresser in the same afternoon, and later tries to recall which person recommended a henna rinse. Even with faded memory traces, schematic processing seems likely to provide the correct answer (see Spaniol & Bayen, 2002, for a review of schemata used in source monitoring). However, if the doctor actually dispenses hair-care tips, the combination of weak episodic traces and strong schemata will lead to a confidently held, false memory. Of course, this example is benign. It would be worrisome, however, if the hairdresser enjoyed giving medical advice.

From a therapeutic or law-enforcement perspective, the question is whether (or to what extent) vivid imagery should be used in probing memory. Studies suggest that imagining is most likely to elicit false memories when people have truly witnessed closely related events (Henkel et al., 2000). Drivdahl and Zaragoza (2001) pressed people to elaborate on suggested events from a crime-scene video; participants later falsely remembered seeing the suggested events. The present results suggest that the same elaborative process used to assist memory in eyewitnesses may actually create memory for unseen actions, especially if the novel ideas are similar to witnessed events. These imagination exercises can also reinforce memories of ‘‘likely suspects’’ as responsible parties, suggesting that witnesses should not engage in vivid imagery as a memory aid. This is especially true if many common circumstances are in play: For example, imagery seems more likely to foster false memories if considerable time will elapse before recollection is required. If multiple people are present during some witnessed event, typical classification processes (e.g., recalling that a young man, two teenage girls, and an older couple were present) will naturally lead to stereotype-consistent recall. The combined influence of vivid imagery and schematic knowledge may cause a store clerk to confidently ‘‘remember’’ a teenage girl stealing a purse, only to express shock later when security tapes show a young man slipping it under his jacket.

Manuscript received 12 March 2007
Manuscript accepted 24 December 2007
First published online 12 February 2008

REFERENCES

Arbuthnott, K. D., Arbuthnott, D. W., & Rossiter, L. (2001). Guided imagery and memory: Implications for psychotherapists. *Journal of Counseling Psychology*, *48*, 123–132.

Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*, 126–145.

Bayen, U. J., Nakamura, B. V., Dupuis, S. D., & Yang, C. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*, 480–500.

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

Chernick, M. R. (1999). *Bootstrap methods: A practitioner's guide*. New York: Wiley-Interscience.

Craik, F., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294.

Drivdahl, S. B., & Zaragoza, M. S. (2001). The role of perceptual elaboration and individual differences in the creation of false memories for suggested events. *Applied Cognitive Psychology*, *15*, 265–281.

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

Garry, M., Frame, S., & Loftus, E. F. (1999). Lie down and let me tell you about your childhood. In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 113–124). New York: Wiley.

Henkel, L. A., & Franklin, N. (1998). Reality monitoring of physically similar and conceptually related objects. *Memory & Cognition*, *26*, 659–673.

Henkel, L. A., Franklin, N., & Johnson, M. K. (2000). Cross-modal source monitoring confusions between perceived and imagined events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 321–335.

Hicks, J. L., & Cockman, D. W., (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language*, *48*, 489–501.

Hicks, J. L., Marsh, R. L., & Ritschel, L. (2002). The role of recollection and partial information source monitoring. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 503–508.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember–know paradigm. *Memory & Cognition*, *25*, 345–351.

Intraub, H., & Hoffman, J.E. (1992). Reading and visual memory: Remembering scenes that were never seen. *American Journal of Psychology*, *105*, 101–114.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, *16*, 337–342.

Johnson, M. K., 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.

Kensinger, E. A., & Schacter, D. L., (2006). Reality monitoring and memory distortion: Effects of negative, arousing content. *Memory & Cognition*, *34*, 251–260.

Kleider, H. M., Pezdek, K., Goldinger, S. D., & Kirk, A. (2008). Schema-driven source misattribution errors: Remembering the expected from a witnessed event. *Applied Cognitive Psychology*, *22*, 1–20.

Lampinen, J. M., Meier, C. R., Arnal, J. D., & Leding, J. K. (2005). Compelling untruths: Content borrowing and vivid false memories. *Journal of Experimental Psychology*, *31*, 954–963.

Lane, S. M., & Zaragoza, M. S. (1995). The recollective experience of cross-modality confusion errors. *Memory & Cognition*, *23*, 607–610.

Lindsay, D. S., & Johnson, M. K. (2001). False memories, fuzzy trace theory, and the source monitoring framework. *Learning and Individual Differences*, *12*, 145–161.

Lyle, K. B., & Johnson, M. K. (2006). Importing perceived features into false memories. *Memory*, *14*, 197–213.

Macrae, C. N., Schloerscheidt, A. M., Bodenhausen, G. V., & Milne, A. B. (2002). Creating memory illusions: Expectancy-based processing and the generation of false memories. *Memory*, *10*, 63–80.

Mammarella, N., & Cornoldi, C. (2002). Aging and the effect of predictability on reality monitoring. *American Journal of Psychology*, *115*, 331–350.

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory*, *14*, 149–160.

Mather, M., & Johnson, M. K. (2003). Affective review and schema reliance on memory in older and younger adults. *American Journal of Psychology*, *2*, 169–189.

Mather, M., Johnson, M. K., & De Leonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology*, *16*, 437–458.

McDaniel, M. A., & Einstein, G. O. (1986). Bizarreness as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 54–65.

Mitchell, K. L., & Johnson, M. K. (2000). Source monitoring: Attribution mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179–195). New York: Oxford University Press.

Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*, 1110–1115.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*, 1–75.

Sherman, J. W., & Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science*, *10*, 106–110.

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 631–651.

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology*, *9*, 101–118.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, *26*, 1–12.