

# Social influences on reality-monitoring decisions

HUNTER G. HOFFMAN

*University of Washington, Seattle, Washington*

PÄR ANDERS GRANHAG

*Göteborg, University, Göteborg, Sweden*

SHEREE T. KWONG SEE

*University of Alberta, Edmonton, Alberta, Canada*

and

ELIZABETH F. LOFTUS

*University of Washington, Seattle, Washington*

A modified Asch (1951) conformity paradigm was used to study the impact of social influence on reality-monitoring decisions about new items. Subjects studied pictures of some objects and imagined others. In a later test phase, they judged whether items had been perceived in the study phase, had been imagined, or were new. Critically, for some items, the subjects were informed of a confederate's response before rendering a judgment. Although the confederate was always correct when they responded to old items, for new items, the confederate responded *perceived*, *imagined*, or *new*, or did not respond (baseline). In two experiments, we show that memory for new items was influenced by an erroneous response of the confederate. Social conformity was reduced by undermining the credibility of the confederate (Experiments 1A and 1B), and the confederate's influence was evident even after there was only a 20-min delay between study and test (Experiment 2), when the subjects were 87% accurate on new baseline items. These experiments reveal the power of social influence on reality-monitoring accuracy and confidence.

Source attribution is a cognitive ability implicated in such memory phenomena as the reliability of eyewitness testimony (Lindsay, 1990; Lindsay & Johnson, 1989; E. F. Loftus & Hoffman, 1989), virtual-reality monitoring (Hoffman, Garcia-Palacios, Thomas, & Botella, 2001), amnesia (Hirst, 1982; Schacter, Harbluck, & McLachlan, 1984), age-related memory deficits (Hashtroudi, Johnson, & Chrosniak, 1989, 1990; McIntyre & Craik, 1987), and the distinction between controlled and automatic memory processes (Jacoby, 1991). Reality monitoring is the decision process by which people separate memories of real events from memories of imagined events, which was proposed by Johnson and Raye (1981; see Johnson, Hashtroudi, & Lindsay, 1993, for a review). Johnson and colleagues proposed that people typically infer the source of origin of their memories at the time of retrieval on the basis of qualities associated with the retrieved memory. The act of self-

generation requires mental effort involving various cognitive operations. Evidence of this effort is preserved in memory. For example, associated memories of false starts at self-generating an antonym might come to mind during the test and serve as cues that the memory was originally imagined. In contrast to memories for imagined events, memories for real events are likely to be rich in cues to their external perceptual origin: colors, perceptual clarity and vividness (sensory properties), contextual information, such as relative spatial location and temporal order (Johnson & Raye, 1981; Johnson, Raye, Foley, & Kim, 1982; see also Johnson, Foley, Suengas, & Raye, 1988), and affect (Johnson et al., 1988; Johnson, Nolde, & De Leonardis, 1996; Johnson & Suengas, 1989). The present study explores a factor besides qualitative characteristics that might be involved in reality-monitoring decisions: social influence. We explored the impact of social influence on the source attribution process.

Social psychologists have studied how information presented to subjects at the time of perception can influence reports and, perhaps, even the phenomenological experience of viewers. In his now famous study, Asch (1951) showed the dramatic influence of conformity on perception tasks. He had control group subjects compare lines of different lengths with a standard line and then had them decide which comparison line matched the standard, a task the subjects were able to perform with almost perfect accuracy. The subjects in an experimental group

---

This research was supported by Mr. and Mrs. George Forbes, by the American-Scandinavian Foundation (to P.A.G.), and by the Social Sciences and Humanities Research Council of Canada (to S.T.K.S.). We thank the reviewers of an earlier version of this paper, originally submitted in 1994, and especially Steve Lindsay and an anonymous reviewer for extensive helpful comments on the present version. Correspondence concerning this article should be addressed to H. G. Hoffman, Human Interface Technology Laboratory, P. O. Box 352142, University of Washington, Seattle, WA 98195-2142 (e-mail: hunter@hitl.washington.edu; <http://www.hitl.washington.edu/people/hunter/>).

made the same perceptual judgment, but only after hearing the responses of numerous confederates. On prearranged trials, each confederate consistently picked a blatantly inaccurate comparison line. The subject "faced, possibly for the first time in his life, a situation in which a group unanimously contradicted the evidence of his senses" (Asch, 1951, p. 179). Faced with this dilemma, over two thirds of the subjects went along with the crowd on at least 1 of the 12 critical trials.

In addition to numerous studies exploring how social influences can affect perception, social psychologists have successfully expanded Asch's (1951) findings to explore how social influence can change the behavior or attitudes of others (e.g., Baron & Byrne, 1994). Despite over a hundred experiments using his classical paradigm to study social influence on behaviors, attitudes, perceptions, and judgments, relatively little research has been directed at exploring how similar social pressure might influence memory. In a notable exception, Betz, Skowronski, and Ostrom (1996) demonstrated how social pressure can influence memory reports, using an Asch-like conformity paradigm. In their studies, subjects first read a story and then completed a recognition memory test of the story facts. Critically, the subjects then received bogus second hand information about 6 other subjects' responses on the same test. Next, the subjects' memory was assessed for this bogus information (a second hand information recognition test) to determine how much of it was stored in memory. Finally, the subjects' memory for the stimulus story was assessed a second time (the story cued-recall test).

In short, Betz et al. (1996) showed that the subjects' responses on the story cued-recall test were significantly influenced by social pressure, as compared with a control condition in which no bogus second hand information had been received. Furthermore, the subjects were more likely to change their responses when their own original memory was poor. If informed that others gave answers that contradicted their answer, the subjects were more prone to change their initial answers in the following story cued-recall test. A follow-up study showed that even when the subjects were explicitly told that the second hand information was bogus, their initial memories were still influenced by this information (i.e., they could not discount the bogus information). Betz et al. demonstrated, in a clear and powerful way, that social pressure (i.e., in their case the influence of a bogus group) can affect an individual's memory. Schneider and Watkins (1996) and Roediger, Meade, and Bergman (in press) came to similar conclusions. Schneider and Watkins had subjects study a word list and found that recognition memory was influenced by a confederate's responses. Roediger et al. showed subjects and confederates common household scenes (e.g., a kitchen). On a collaborative recall task, memory for objects in the scenes was influenced by the confederate's responses.

As has been noted by Nelson, Kruglanski, and Jost (1998), figuring out what other people know (social metacognition) is essential for successful social interaction. They argued that such metacognitive assessments of others' knowledge is important for interpersonal communication, social influence (e.g., accurately predicting what lines of argument others may find persuasive), giving social support (understanding the other person's needs), and deciding which members of a group are the most valuable source of information. The research reported here extends the findings of Betz et al. (1996), Schneider and Watkins (1996), and Roediger et al. (in press), by exploring the impact of social influence on metamemory processes.

In the present study, subjects studied pictures of some objects and imagined others. In a later test phase, the subjects were shown the name of an object and indicated whether it had earlier been seen (perceived) or imagined or whether the object had not been presented earlier (new). Prior to each test item (except for baseline items), the subjects were given access to the responses of a "previous subject" (actually a confederate). The confederate was always correct when he or she responded to old test items. On new trials, the confederate's response was congruent with the correct answer (e.g., confederate responded *new*, and the item was new), was incongruent (e.g., confederate responded *imagined* when the item was new), or was a baseline trial (e.g., confederate said nothing, and the item was new). Of interest were subjects' memory reports for new items. Showing the influence of conformity, we expected that memory judgments about new items would be compromised by the confederates' erroneous responses (i.e., the subject would be more likely to report that a new item was imagined if the confederate said it was imagined). On ratings of confidence, we expected that responding correctly even in the face of social pressure would reduce confidence, whereas subjects would be more confident in their incorrect responses when they went along with an erroneous confederate response. In addition to the trial type manipulation described above, in Experiments 1A and 1B the credibility of the confederate was manipulated. Susceptibility to misinformation in the eyewitness context has been found to depend on the perceived credibility of the source of misinformation. For example, Smith and Ellsworth (1987) found that if the source of erroneous postevent information is thought to be competent and knowledgeable, this person's erroneous information is believed, and memory accuracy is compromised (see also Kwong See, Hoffman, & Wood, in press). Similarly, Dodd and Bradshaw (1980) found that if subjects were led to believe the source was not honest because the source had an intention to mislead, erroneous information provided by this source did not influence memory reports. In Experiments 1A and 1B of the present study, either the confederate was described as a credible person or the subject was told that the response was randomly generated by a computer

and, thus, the response was not trustworthy. The goal was to determine whether conformity would be lessened when the confederate had low credibility.

As was mentioned earlier, Asch (1951) was able to show social influences on perception when accuracy on the perception task would otherwise have been nearly 100%. In Experiment 2 of the present study, the delay between study and test was decreased to 20 min to determine whether subjects would still conform, even in the 20-min delay group, when their memories for baseline trials (i.e., when there was no confederate) should be higher than that for a 1-day delay group.

## EXPERIMENT 1A

### Method

**Subjects.** Forty-six undergraduate students from the University of Washington participated for extra credit in an introductory psychology course.

**Materials.** Seventy-two objects were chosen from category exemplars described by Battig and Montague (1969). Nine exemplars from each of the following eight categories were used: fruits, birds, land animals, tools, kitchen items, vehicles, musical instruments, and sea creatures. Two slides of each object were adapted from Hoffman (1997). One slide was a color photo of the object on a white background (perceived items), and the other was blank (used for imagined items, described below).

The reality-monitoring test consisted of a stack of seventy-two, 3 × 4-in. squares of paper. The confederate's response was printed and circled at the top of the test card. The test item was printed in the middle (center) of the test card. Below the test item were the response choices: perceived, imagined, or new. On the line below was a confidence rating scale (0%, 20%, 40%, 60%, 80%, or 100%).

**Design and Procedure.** A 2 × 3 mixed experimental design was employed, with instruction (high credibility vs. low credibility) as a between-subjects factor and trial type (congruent, incongruent, or baseline) as a repeated measure. The experiment was described to the subjects as an investigation of the psychological processes involved in perceiving and imagining. Each subject viewed a slide tray of alternating pictures and blank screens. Every 5 sec, they heard the name of an object via a tape recorder. Sometimes the name was followed immediately by a picture of the object. Other times, the name was followed by a blank screen, and the subjects were to create a mental image of the object. As a cover task for the latter, the subjects were to judge how long it would take an artist to make a line drawing of the object. The subjects recorded their judgments on a scoring sheet by circling numbers ranging from 1 (for *very little time to draw*) to 7 (for *a lot of time to draw*). No mention was made of a subsequent test phase.

Twelve perceived items and 12 imagined items were presented to the subjects in the study phase according to a counterbalancing scheme described below. To decrease retrieval accuracy for all the subjects for these critical study items, the items were preceded by 48 filler trials. The filler trials comprised 24 perceived items and 24 imagined items. The same filler items were shown in the same order to all the subjects. Memory for the filler items was never tested.

After the study phase, the subjects completed a 20-min distractor task (completing word puzzles). The participants returned for the memory test after a 48-h delay. The reality-monitoring test included the 24 critical items studied earlier and 48 new test items not encountered before in the experiment. The order of item presentation in the test was randomized, with the restriction that no more than 3 items of the same class (e.g., imagined items) could be presented in a row. The items from the reality-monitoring test were counterbalanced as follows. The 72 exemplars in the test item pool

were divided into six subsets of 12 items each. The mean ranking of item prototypicality (derived by Battig & Montague, 1969) was equated for each subset. The stimuli were rotated through the six conditions listed below. In the course of the experiment, each stimulus appeared in each of the six within-subjects experimental conditions approximately equally often. The conditions for the new items were the following: (1) confederate responds *perceived*, test item is new (incongruent); (2) confederate responds *imagined*, test item is new (incongruent); (3) confederate responds *new*, test item is new (congruent), and (4) confederate does not respond, test item is new (baseline). The conditions for the old items were the following: (1) confederate responds *perceived*, test item was perceived at study; and (2) confederate responds *imagined*, test item was imagined at study.

Baseline trials were designed to measure the extent to which the subjects were relying on memory for the stimuli when no confederate's response preceded the test item. Across subjects, each item was used as a baseline trial approximately equally often.

All the subjects were given a stack of test items and an instruction sheet. They were reminded that earlier they had listened to a list of items, some of which had been accompanied by pictures and some of which they had imagined seeing. They were told that they would now be given a memory test for these pictures and imagined objects and that new items absent from the study phase had been added to the test. For each test item, they were to identify (by circling) whether the object had been perceived or imagined or whether the item was new and then to indicate how confident they were in the accuracy of their response as a percentage—0% (*completely guessing*), 20%, 40%, 60%, 80%, or 100% (*completely sure*).

The subjects were randomly assigned to one of two groups. The only difference between these groups was whether they received *high-credibility* instructions or *low-credibility* instructions. The subjects in the high-credibility group were told that each test item would have the response of a previous subject accompanying it on the test card. The other respondent, they were told, was a graduate student from the University of Washington, who had studied the same items and had taken a memory test similar to the test the subjects were about to take. The subjects were instructed that they were free to go along with the graduate student's response or to ignore her response.

In the low-credibility condition, the subjects were told that the other response they saw on the test card was randomly generated by a computer, the computer equivalent of flipping a coin. In both groups, the subjects were asked to make a mark next to this response to indicate that they had looked at it and then to make their own memory decisions and confidence ratings. In all the conditions described in this paper, the subjects were encouraged to make their source identifications fairly quickly, because they would be given only a few seconds per object to make their decisions. In fact, no time limit was imposed.

### Results

All *p* values reported are two-tailed. An alpha level of .05 was used for all the statistical tests reported in this paper.

**Accuracy of responses to new items.** We first calculated the percentage of times that the subjects accurately responded *new* to a new item and the percentage of inaccurate responses to new items in the high-credibility condition. These data are shown in Figure 1. This figure shows the percentage of *new*, *perceived*, and *imagined* responses given by the subjects as a function of the confederate's response. Keep in mind that for the data in this figure, the correct response was always *new*. For ease of exposition, throughout this manuscript, *baseline* refers to new items where the confederate said nothing, and *old*

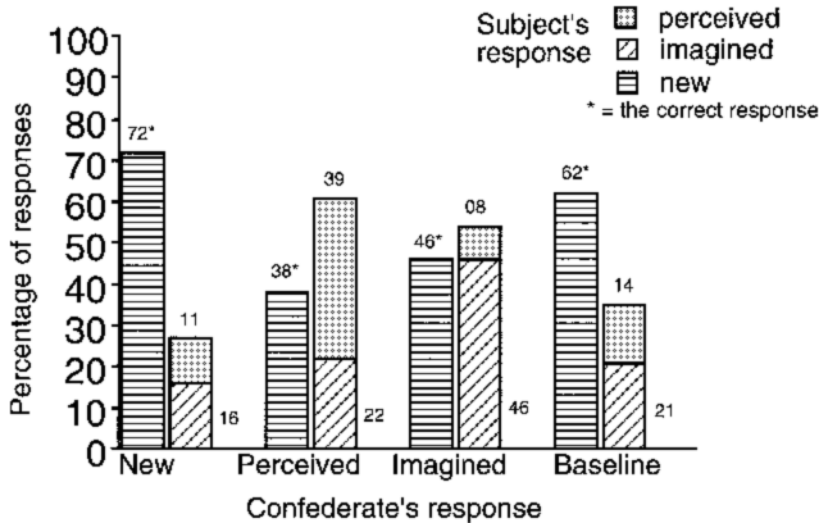


Figure 1. Percentages of responses in the high-credibility group of Experiment 1A, 2 days after the study phase.

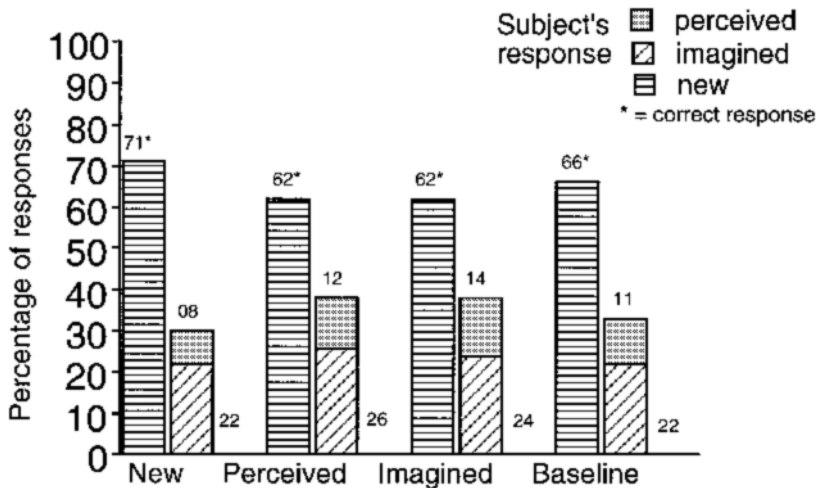


Figure 2. Percentage of responses in the low-credibility group of Experiment 1A, 2 days after the study phase.

*baseline* refers to old items where the confederate said nothing (used only in Experiment 1B).

Statistical tests support what is readily apparent in Figures 1 and 2. A 2 (confederate credibility: high vs. low)  $\times$  3 (item type: congruent, incongruent, or baseline) analysis of variance (ANOVA) was conducted. No significant between-subjects main effect occurred for confederate credibility [ $F(1,44) = 2.34$ ,  $MS_e = 0.18$ ,  $p = .13$ ], although there was a trend in the predicted direction. A main effect was found for item type [ $F(2,88) = 15.65$ ,  $MS_e = 0.44$ ,  $p < .001$ ]. An interaction between confederate credibility and item type was found [ $F(2,88) = 4.82$ ,  $MS_e = 0.14$ ,  $p = .01$ ]. When a low-credibility confederate

responded inaccurately (i.e., responded either *perceived* or *imagined* to a new item), the subjects' performance was not deleteriously affected, as compared with the baseline [62% vs. 66%, respectively;  $t(22) = 1.12$ ,  $p > .10$ , n.s.]. When a low-credibility confederate responded accurately, the subjects' performance was not increased significantly, as compared with the baseline [71% vs. 66%, respectively;  $t(22) < 1$ , n.s.].

In contrast, when a high-credibility confederate responded inaccurately, performance was dramatically affected, as compared with the baseline. Performance dropped substantially from 62% to 42% accuracy [ $t(22) = 4.80$ ,  $SE = 0.04$ ,  $p < .001$ ]. A nonsignificant trend to-

ward higher performance when the confederate responded accurately was also evident [72% vs. 62%;  $t(22) = 1.56, SE = 0.06, p = .13$ ].

As is shown in Figure 1 regarding baseline items for the high-credibility condition, when the subjects erred, they were more likely to respond *imagined* than *perceived* [ $t(22) = 2.13, SE = 0.03, p < .05$ ]. The same pattern was found for the low-credibility condition [ $t(22) = 2.65, SE = 0.04, p < .05$ ].

**Confidence in correct responses to new items.** An analysis using the same mixed design showed no significant between-groups difference in overall level of confidence on correct responses [i.e., high vs. low credibility;  $F(1,42) < 1, n.s.$ ]. The within-subjects factor (item type: congruent, incongruent, or baseline) did have a significant effect on confidence [ $F(2,84) = 16.07, MS_e = 119.00, p < .001$ ]. Collapsed across groups, confidence on congruent trials was higher than confidence on incongruent trials [75% vs. 62%;  $t(43) = 4.80, SE = 2.65, p < .001$ ]. Confidence on congruent trials was higher than on baseline trials (75% vs. 70%;  $t(44) = 2.31, SE = 2.04, p < .05$ ), and confidence on incongruent trials was significantly lower than on baseline trials [62% vs. 70%;  $t(44) = 3.69, SE = 2.37, p = .001$ ].

There was a marginally significant interaction between the high- and the low-credibility groups with respect to confidence [ $F(2,84) = 2.73, MS_e = 324.61, p = .07$ ]. For the low-credibility group, confidence on congruent trials was not significantly different from baseline [72% vs. 68%;  $t(22) = 1.48, p > .10, n.s.$ ]. And confidence on incongruent trials did not differ significantly from baseline [64% vs. 68%, respectively;  $t(22) = 1.65, p > .10, n.s.$ ].

For the high-credibility group, confidence was adversely affected by the erroneous confederate. That is, confidence ratings were significantly lower on incongruent trials than on baseline trials [60% vs. 72%, respectively;

$t(21) = 3.51, SE = 3.88, p < .005$ ]. And confidence ratings were higher on congruent trials than on baseline trials (77% vs. 72%, respectively), and this difference was marginally significant [ $t(21) = 1.75, p < .10$ ].

Showing concordance between ratings of confidence and accuracy, the subjects in the low-credibility group gave realistic confidence judgments for all three types of trials. In contrast, the subjects in the high-credibility group were highly overconfident for both incongruent and baseline trials (i.e., they thought they had made correct decisions more often than proved to be the case).

**Accuracy of responses to old items.** A large number of test trials involving old items were included primarily to help establish the confederate's credibility with the subject. That is, the confederate was always correct on the old items, never incorrect. Although a less convincing demonstration of the impact of conformity than when new items are considered, it is possible to explore the impact of the confederate on old-item performance. Because we found that the low-credibility ("random") confederate had virtually no impact on new-item performance, we expected that the low-credibility confederate would have no effect on old-item performance. In the high-credibility condition, we expected an impact for the old items, as had been shown for the new items. The means of the data used in these between-groups calculations are shown in Figure 3.

An examination of Figure 3 reveals that when the high-credibility confederate accurately responded *perceived*, this increased the likelihood that the subject would accurately respond *perceived* [84%, as compared with 67% for the low-credibility group];  $t(44) = 3.64, SE = 0.05, p = .001$ , but made no difference in confidence [88% in the high-credibility group, as compared with 90% confidence for the low-credibility group;  $t(43) < 1, n.s.$ ]. When the high-credibility confederate accurately re-

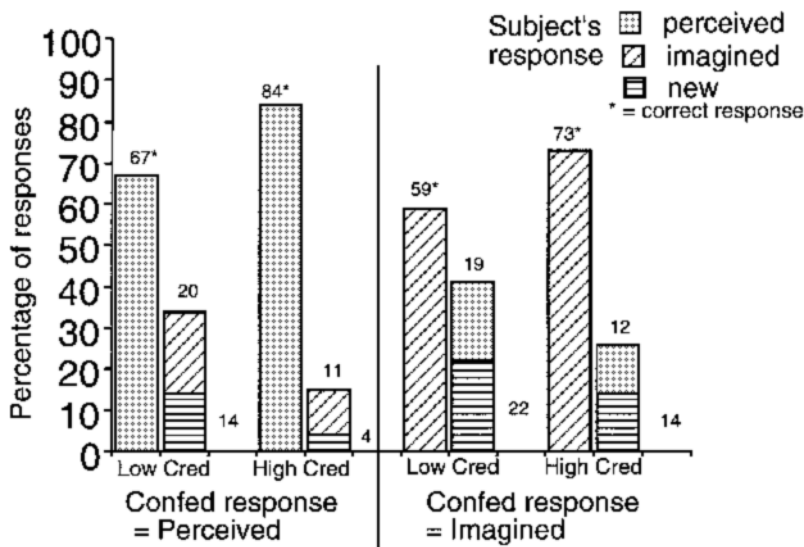


Figure 3. Source identifications on old items in Experiment 1A.

sponded *imagined*, this increased the likelihood that the subject would accurately respond *imagined* [73%, as compared with 59% for the low-credibility group;  $t(44) = 2.26$ ,  $SE = 0.06$ ,  $p < .05$ ] but made no difference in confidence [78% for the high-credibility group, as compared with 77% for the low-credibility group;  $t(43) < 1$ , n.s.]. Thus, the confederate also boosted the likelihood of an accurate response on old items.

**Source attribution biases on old items.** As is shown on the left side of Figure 3, collapsing across high and low confederate credibility, when the subjects erred on an old perceived item, they were more likely to respond *imagined* than *new* [15% vs. 9%, respectively;  $t(45) = 2.04$ ,  $SE = 0.03$ ,  $p < .05$ ]. In contrast, when the subjects erred on an old imagined item, they were not more likely to respond *perceived* than *new* [15% vs. 18%, respectively;  $t(45) < 1$ , n.s.].

## EXPERIMENT 1B

There is one aspect of the design of Experiment 1A that could conceivably have influenced responding in a rather uninteresting way. Every time no confederate response was given, the correct response was *new*. If the subject even inadvertently picked up on this association, this could have elevated baseline performance, artificially exaggerating the impairment in performance obtained in the face of an erroneous confederate. To ensure that this design feature was not responsible for the impact of the confederate, Experiment 1B was conducted. The correct response for baseline items was equally likely to be *old* or *new*.

### Method

**Subjects.** Fifty undergraduate students from the University of Alberta, Canada participated for extra credit in an introductory psychology course.

**Materials.** Eighty-four objects adapted from Hoffman (1997) were used: The 72 items from Experiment 1A, plus 12 additional objects from the same stimulus pool.

**Design and Procedure.** The design, instructions, and study phase were identical to those used in Experiment 1A. The items for the reality-monitoring test were counterbalanced as follows. The 84 exemplars in the test item pool were divided into seven subsets of 12 items each, the six subsets used in Experiment 1A, and one subset not used in Experiment 1A—confederate does not respond, test item is old (baseline). The stimuli were rotated through the seven conditions. In the course of the experiment, each stimulus appeared in each of the seven within-subjects experimental conditions approximately equally often. Other than the small change in stimuli mentioned above (the addition of old baseline items), the procedure was identical to that used in Experiment 1A.

### Results

**Accuracy of responses to new items.** The results of Experiment 1B showed the same overall pattern as did those of Experiment 1A. Findings that were not consistent and/or not consistently significant across Experiments 1A and 1B are explicitly highlighted in the Exper-

iment 1B Discussion section. No figures are shown for Experiment 1B. We first calculated the percentage of times that the subjects accurately responded *new* to a new item and the percentage of inaccurate responses to new items in the high-credibility condition.

A 2 (confederate credibility: high vs. low)  $\times$  3 (item type: congruent, incongruent, or baseline) ANOVA was conducted. A between-subjects main effect was found for confederate credibility [ $F(1,48) = 5.11$ ,  $MS_e = 0.34$ ,  $p < .05$ ]. A main effect was found for item type [ $F(2,96) = 28.72$ ,  $MS_e = 0.79$ ,  $p < .001$ ]. An interaction between confederate credibility and item type was found [ $F(2,96) = 6.62$ ,  $MS_e = 0.18$ ,  $p < .005$ ]. When a low-credibility confederate responded inaccurately (i.e., responded either perceived or imagined to a new item), the subjects' performance was not deleteriously affected, as compared with the baseline [63% incongruent items vs. 68% for baseline;  $t(24) = 1.07$ ,  $p > .05$ , n.s.]. And when a low-credibility confederate responded accurately, the subjects' performance showed only a marginally significant increase in accuracy, as compared with the baseline, [76% for congruent vs. 68% for baseline;  $t(24) = 1.79$ ,  $SE = 0.05$ ,  $p = .09$ ].

In contrast, when a high-credibility confederate responded inaccurately, performance was dramatically affected, as compared with baseline (new items where the confederate said nothing). Performance dropped substantially from 59% to 42% in accuracy [ $t(24) = 4.15$ ,  $SE = 0.04$ ,  $p < .001$ ]. When a high-credibility confederate responded accurately, the subjects' performance was increased significantly, as compared with baseline [59% baseline vs. 79% for congruent items;  $t(24) = 4.67$ ,  $SE = 0.04$ ,  $p < .001$ ].

Regarding baseline items for the high-credibility condition, when the subjects erred, they were more likely to respond *imagined* than *perceived* [31% vs. 12%, respectively;  $t(23) = 6.04$ ,  $SE = 0.03$ ,  $p < .001$ ]. The same pattern was found for the low-credibility condition (19% for *imagined* vs. 13% for *perceived*) but was only marginally significant [ $t(24) = 1.84$ ,  $SE = 0.03$ ,  $p = .08$ , n.s.].

**Confidence in correct responses to new items.** Analyses of confidence ratings, using the same mixed design, showed no significant between-groups difference in overall level of confidence [i.e., high vs. low credibility;  $F(1,44) < 1$ , n.s.]. The within-subjects factor [item type: congruent, incongruent, or baseline] did have a significant effect on confidence [ $F(2,88) = 6.72$ ,  $MS_e = 542.20$ ,  $p < .005$ ]. Collapsed across groups, confidence on congruent trials was marginally higher than baseline [71% vs. 68%;  $t(49) = 1.71$ ,  $SE = 1.77$ ,  $p = .095$ ], and confidence on incongruent trials was significantly lower than baseline [64% vs. 68%;  $t(45) = 2.71$ ,  $SE = 1.52$ ,  $p = .009$ ].

There was no significant interaction between the high- and the low-credibility groups with respect to confidence [ $F(2,88) = 2.12$ ,  $p > .05$ ]. For the low-credibility group, confidence on congruent trials was not significantly different from baseline [70% vs. 68%, respectively;  $t(24) <$

1, n.s.]. And confidence on incongruent trials did not differ significantly from baseline [68% vs. 68%, respectively;  $t(23) < 1$ , n.s.].

For the high-credibility group, confidence was adversely affected by the erroneous confederate. That is, confidence ratings were significantly lower on incongruent trials than on baseline trials [61% vs. 68%, respectively;  $t(21) = 3.42$ ,  $SE = 2.21$ ,  $p = .003$ ]. But confidence was not boosted by an accurate confederate. Ratings were higher on congruent trials than on baseline trials (71% vs. 67%, respectively), but this difference was not statistically significant [ $t(24) = 1.55$ ,  $p > .05$ ].

Showing concordance between ratings of confidence and accuracy, the subjects in the low-credibility group gave realistic confidence judgments for all three types of trials. In contrast, the subjects in the high-credibility group were highly overconfident for both incongruent and baseline trials (i.e., they thought they had made correct decisions more often than proved to be the case).

**Accuracy of responses to old items.** A 2 (confederate credibility: high vs. low)  $\times$  2 (item type: congruent or baseline) ANOVA was conducted. No significant between-groups main effect occurred for confederate credibility [ $F(1,48) < 1$ , n.s.]. A main effect was found for item type [ $F(1,48) = 19.26$ ,  $MS_e = 0.32$ ,  $p < .001$ ]. Collapsing across groups, a significant difference was found comparing accuracy on congruent trials (.67) with that on baseline trials [.56;  $t(49) = 4.57$ ,  $SE = 0.03$ ,  $p < .001$ ]. We next compared accuracy on congruent old perceived trials (.70) with accuracy on baseline old perceived trials (.56). Collapsing across groups, a significant difference was found for accuracy [ $t(49) = 3.98$ ,  $SE = 0.04$ ,  $p < .001$ ]. We next compared accuracy on congruent old imagined trials (.65) with accuracy on baseline old imagined trials (.55). Collapsing across groups, a significant difference was found for accuracy [ $t(49) = 2.72$ ,  $SE = 0.04$ ,  $p < .01$ ]. No interaction between confederate credibility and item type was found [ $F(1,48) < 1$ , n.s.].

**Confidence in correct responses to old items.** A 2 (confederate credibility: high vs. low)  $\times$  2 (item type: congruent or baseline) ANOVA was conducted. No significant between-groups main effect occurred for confederate credibility [ $F(1,48) < 1$ , n.s.]. No main effect occurred for item type [ $F(1,46) = 1.06$ ,  $p > .10$ , n.s.]. No interaction between confederate credibility and item type was found [ $F(1,46) < 1$ , n.s.].

**Source attribution biases on old items.** Collapsing across high and low confederate credibility, when the subjects erred on an old perceived item, they were more likely to respond *imagined* than *new* [21% vs. 9%, respectively;  $t(49) = 3.84$ ,  $SE = 0.03$ ,  $p < .001$ ]. In contrast, when the subjects erred on an old imagined item, they were not more likely to respond *perceived* than *new* [16% vs. 19%, respectively;  $t(49) < 1$ , n.s.].

## Discussion

The results of Experiments 1A and 1B provide strong evidence that conformity influences reality-monitoring decisions. When the high-credibility confederate gave an

inaccurate response to a new item (incongruent trials), accuracy dropped substantially, as compared with the baseline, in both experiments. Moreover, when the subjects resisted the confederate's influence and responded accurately on incongruent trials, confidence was reduced in both experiments (as compared with confidence on baseline items). The effect was less pronounced for congruent trials. When a high-credibility confederate accurately responded *new* to a new item (congruent trials), the subjects conformed to that accurate response (a nonsignificant trend in Experiment 1A and a significant increase in accuracy on congruent vs. baseline trials in Experiment 1B). On such congruent trials with a high-credibility confederate, confidence levels rose marginally in Experiment 1A but did not increase significantly in Experiment 1B.

The results of Experiments 1A and 1B also nicely show that the perceived credibility of the confederate mediates the impact of social conformity. When the subjects believed that the confederate's responses were randomly generated by a computer and thus not deliberate (low confederate credibility condition), in both Experiments 1A and 1B, the pattern of responses to new items indicated that the low-credibility confederate's responses were essentially ignored on incongruent trials. The subjects' accuracy on incongruent trials was not significantly different from accuracy on baseline trials in either Experiment 1A or 1B. On congruent trials, the low-credibility confederate did not significantly influence accuracy in Experiment 1A (as compared with baseline items) and had only a marginally significant influence on accuracy in Experiment 1B. And confidence was not affected by congruent or incongruent confederates in either experiment.

Social influence also influenced reality-monitoring decisions on old items. An accurate confederate enhanced the accuracy of the subjects' responses to old items in both Experiments 1A and 1B. And this was true for both perceived items and imagined items (i.e., both showed facilitation on congruent trials). When occasionally errors were made on old perceived items, the subjects in both experiments showed a bias to call it *imagined* rather than *new*. On the other hand, when the subjects erred on an imagined item, they exhibited no tendency to call the item *perceived* rather than *new*, and this was the case in both experiments.

## EXPERIMENT 2

In Experiment 1, there was a 48-h delay between study and test. This delay may have contributed to the relatively low baseline performance—64% correct on the new items in Experiments 1A and 1B. With a short retention interval, baseline performance would be expected to be higher, since memory for actual items would be stronger and subjects should more readily distinguish items they had seen from ones that they had not. In Experiment 2, the impact of the confederate was examined when memory for test items was stronger. Perhaps the impact of the confederate would be substantially lessened in this case. Supporting this supposition, there is evidence that when

initial memory is good rather than poor, subjects are less readily persuaded by misinformation (E. F. Loftus, Miller, & Burns, 1978). Thus, Experiment 2 included a 20-min retention interval and compared the resulting data with those obtained with a longer (24-h) retention interval.

**Method**

**Subjects.** The subjects were 41 undergraduate students from the University of Washington, who volunteered to participate for extra credit in an introductory psychology course.

**Materials.** Materials were the same as those used in Experiment 1A. (Experiment 2 was conducted before the improved Experiment 1B introduced old baseline items. Experiment 2 does not include old baseline items.)

**Design and Procedure.** A 2 × 3 mixed experimental design was employed, with delay (short vs. long) as a between-subjects vari-

able and trial type (congruent, incongruent, or baseline) as a repeated measure. The subjects were randomly assigned to either the short-delay group (20 min between study and test) or the long-delay group (24 h between study and test). Twenty subjects were in the short-delay group, and 21 were in the long-delay group. All the subjects were given the high confederate credibility instructions from Experiment 1.

**Results**

**Accuracy of responses to new items.** Not surprisingly, the delay of the test did influence performance. Consider what happens when the confederate made no response to new items (baseline trials). As is shown in Figures 4 and 5, baseline accuracy was 69% when the subjects were tested after a 24-h delay and 87% when they were tested after a 20-min delay. For ease of exposition, we separately examine the data for the two delays.

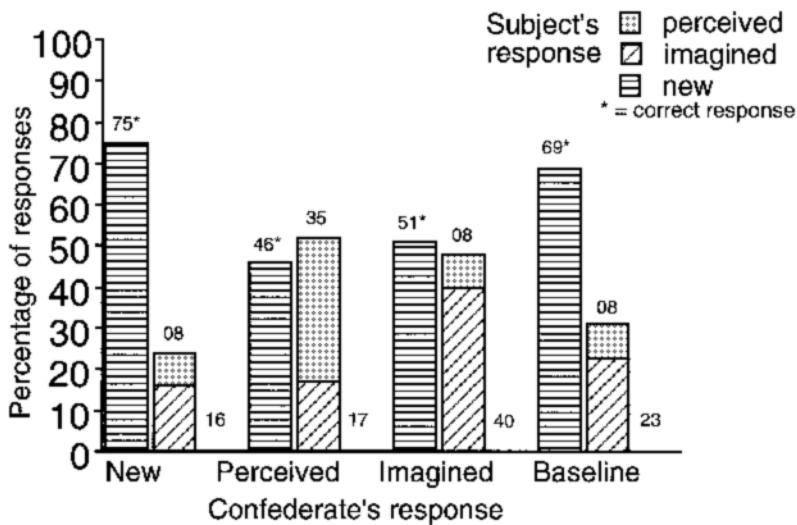


Figure 4. Percentage of responses in the 24-h delay, high-credibility group.

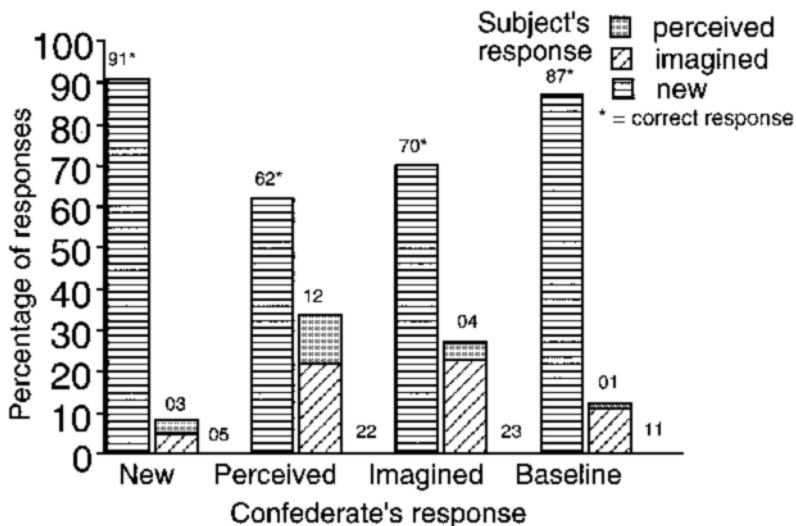


Figure 5. Percentage of responses in the 20-min delay, high-credibility group.



We first calculated the percentage of times that the subjects accurately responded *new* to a new item and the percentage of inaccurate responses to new items in the 24-h delay condition. The latter is the condition most comparable with that used in Experiment 1A. These data are shown in Figure 4. This figure shows the percentage of *new*, *perceived*, and *imagined* responses given by the subjects as a function of the confederate's response. Keep in mind that, for the data in this figure, the correct response was always *new*.

Statistical tests show a significant influence of the length of the delay between study and test on memory accuracy, as well as a powerful influence of the confederate on accuracy. A 2 (delay interval: 24 h vs. 20 min)  $\times$  3 item type: congruent, incongruent, or baseline) mixed model ANOVA was conducted. A between-subjects main effect was found for delay [ $F(1,39) = 17.12, MS_e = 0.87, p < .001$ ]. Delay had a significant effect on baseline accuracy [ $t(39) = 3.75, SE = 0.05, p = .001$ ]. A main effect was found for item type [ $F(2,78) = 28.01, MS_e = 0.77, p < .001$ ]. Collapsing across groups, no significant difference was found when accuracy on congruent trials was compared with that on baseline trials [ $t(40) = 1.62, p > .10, n.s.$ ]. Accuracy on incongruent trials was significantly lower than performance on baseline trials [ $t(40) = 5.42, SE = 0.04, p < .001$ ]. Surprisingly, no interaction between delay and confederate's response was found [ $F(2,78) < 1, n.s.$ ].

**Confidence in correct responses to new items.** Using a mixed ANOVA, no significant between-groups differences in overall level of confidence (i.e., 20-min vs. 24-h delay) was found [ $F(2,74) < 1, n.s.$ ]. A main effect was found for item type [congruent vs. incongruent vs. baseline;  $F(2,36) = 14.52, MS_e = 68.42, p < .001$ ]. Collapsed across groups, confidence on congruent trials was higher than confidence on baseline trials (82% vs. 78%),

but the difference was not significant [ $t(40) = 1.5, SE = 2.09, p = .14$ ]. Confidence on incongruent trials was significantly lower than baseline [72% vs. 78%;  $t(38) = 3.29, SE = 1.90, p < .005$ ]. No interaction between item type and delay was found [ $F(2,36) < 1, n.s.$ ].

**Source misattributions.** With regard to baseline items for the 24-h delay condition, when the subjects erred, they were more likely to respond *imagined* than *perceived* [ $t(20) = 5.23, SE = 0.03, p < .001$ ]. The same pattern was found for the 20-min delay condition [ $t(19) = 4.15, SE = 0.02, p < .001$ ].

**Accuracy of responses to old items.** In Figure 6, we present the old-item data.

Collapsing across short and long delays, when the subjects erred on an old perceived item, they were slightly but not significantly more likely to respond *imagined* than *new* [.12 vs. .08;  $t(40) = 1.67, SE = 0.02, p = .10$ ]. In contrast, when the subjects erred on an old imagined item, they were not more likely to respond *perceived* than *new* [.08 vs. .11, respectively;  $t(40) < 1, n.s.$ ].

## Discussion

Experiment 2 again revealed the power of the confederate to influence both reality-monitoring accuracy and confidence on incongruent trials. Overall, reality-monitoring accuracy was significantly lower when the confederate responded inaccurately than when the confederate did not respond at all (baseline). The study-test delay manipulation was successful. The subjects in the 20-min group showed higher memory accuracy on baseline items than did the subjects in the 24-h group. Interestingly, contrary to what might be predicted on the basis of the social literature, the results of Experiment 2 showed no evidence that social influence was stronger for the 24-h delay group. However, null effects of the sort found in Experiment 2 should be interpreted with caution. We cannot conclude

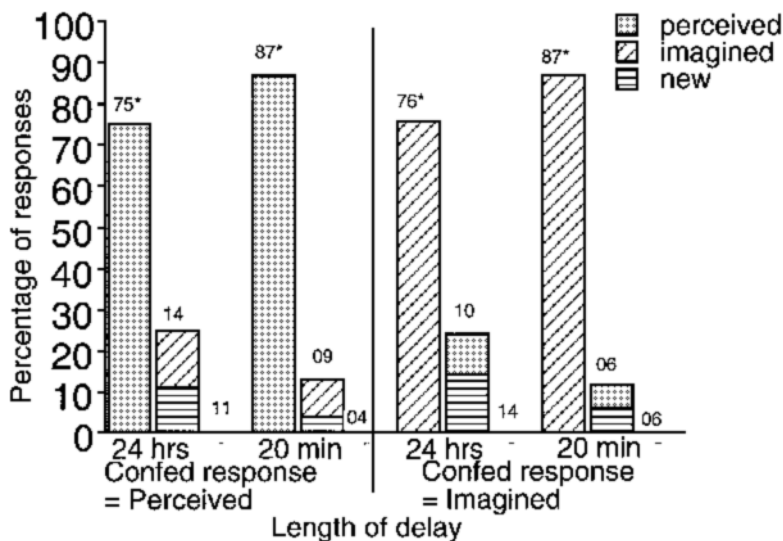


Figure 6. Source identifications on old items in Experiment 2.

that conformity was the same in the two groups on the basis of the results presented here (G. R. Loftus & E. F. Loftus, 1988, p. 162).

## GENERAL DISCUSSION

Johnson and colleagues (see Johnson et al., 1993, for a review) have amassed an impressive literature supporting the notion that people infer the source of origin of their memories at time of retrieval on the basis of qualities associated with the retrieved memory. The present results demonstrate a new factor that contributes to reality-monitoring decisions: social influence from a perceived authority figure. Experiments 1A and 1B showed the impact of erroneous social influence on reality-monitoring decisions and showed that undermining the credibility of the confederate reduced conformity.

Experiment 2 showed that erroneous social influence still deleteriously affected source identification accuracy, even when there was only a 20-min delay between study and test, conditions in which subjects' memories were quite good (87% accurate when no confederate responded). The fact that the delay manipulation successfully affected memory accuracy but did not affect the amount of conformity appears to contradict both the social literature and predictions based on the misinformation paradigm. Although their memories were more accurate, the subjects in the short-delay group were not more confident in the accuracy of their memories than the subjects in the long-delay group. This may help explain why there was no significant change in susceptibility to social influence (i.e., no difference in conformity for the short-delay vs. the long-delay group). According to Strack and Bless (1994; Bless & Strack, 1998), a manipulation (such as a delay) that affects memory accuracy will not necessarily change how much people conform. They proposed that a change in the subject's impression about how accurately he or she can remember items is what determines whether there will be a change in his or her susceptibility to social influence. Using the misinformation paradigm (E. F. Loftus, 1979), Strack and Bless found that subjects in one group (the low salience stimuli group) were less confident in the accuracy of their memories and conformed more, even when the actual memory accuracy performance was the same as that of the other group (the high stimulus salience group).

Overall, the findings of the present study have possible implications for other research topics involving source attribution. Gudjonsson and colleagues (e.g., Gudjonsson, 1986, 1992) have demonstrated that it is possible to change unwanted (and still possibly accurate) parts of a memory report by having an interrogator challenge and give negative feedback to subjects' initial answers. This so-called interrogative suggestibility (Gudjonsson & Clark, 1986) is an example of a type of obtrusive social pressure. Positive feedback from police to eyewitnesses can also bias the witnesses. In a mock lineup identification study, Wells and Bradfield (1999) told eyewitnesses they had accurately identified the perpetrator when, in

fact, the witness had made a false identification. They found that giving the eyewitness erroneous feedback after they made their lineup identifications inflated the witnesses' later recollection of how confident they had been at the time of the identification. Similarly, Kassir and Kiechel (1996) found that presenting people with false incriminating evidence can lead them to confess to an act they did not commit and to internalize the confession and, perhaps, confabulate details in memory.

Whether conforming to the confederate (or interrogator) changes what subjects subsequently remember is an important issue. If this is so, it would involve a transfer from external (social) to internal (memory) sources of influence on performance. Alternatively, instead of distorting memory, social influence may merely alter subjects' reports of what they remember. In the classic conformity literature, a distinction is made between informational and normative influences of conformity (Deutsch & Gerard, 1955). Informational influences are shown when the responses of others are used as a source of information about reality, and conformity reflects the belief that others are correct. In contrast, normative influences are shown by people's conforming even when they know others are wrong, so as to reap the rewards gained by compliance and to avoid the punishments that disagreement brings. Campbell, Tesser, and Fahey (1986) suggest that informational influences are most important when stimuli are highly ambiguous (difficult to judge). Thus, if conformity distorts memory, such distortion is most likely to be evident when the memories are difficult to remember, as in the present studies.

## REFERENCES

- ASCH, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men: Research in human relations* (pp. 177-190). Pittsburgh: Carnegie Press.
- BARON, R. B., & BYRNE, D. (1994). *Social psychology*. New York: Prentice-Hall.
- BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, **80** (3, Pt. 2), 1-46.
- BETZ, A. L., SKOWRONSKI, J. J., & OSTROM, T. M. (1996). Shared realities: Social influence and stimulus memory. *Social Cognition*, **14**, 113-140.
- BLESS, H., & STRACK, F. (1998). Social influence on memory. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 90-106). Thousand Oaks, CA: Sage.
- CAMPBELL, J. D., TESSER, A., & FAHEY, P. J. (1986). Conformity and attention to the stimulus: Some temporal and contextual dynamics. *Journal of Personality & Social Psychology*, **51**, 315-324.
- DEUTSCH, M., & GERARD, H. G. (1955). A study of normative and informational social influence upon individual judgement. *Journal of Abnormal & Social Psychology*, **59**, 204-209.
- DODD, D. H., & BRADSHAW, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning & Verbal Behavior*, **19**, 695-704.
- GUDJONSSON, G. H. (1986). The relationship between interrogative suggestibility and acquiescence: Empirical findings and theoretical implications. *Personality & Individual Differences*, **7**, 195-199.
- GUDJONSSON, G. H. (1992). *The psychology of interrogations, confessions and testimony*. Chichester, U.K.: Wiley.
- GUDJONSSON, G. H., & CLARK, N. K. (1986). Suggestibility in police interrogation: A social psychological model. *Social Behaviour*, **1**, 83-104.

- HASHTROUDI, S., JOHNSON, M. K., & CHROSNIAK, L. D. (1989). Aging and source monitoring. *Psychology & Aging*, **4**, 106-112.
- HASHTROUDI, S., JOHNSON, M. K., & CHROSNIAK, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology & Aging*, **5**, 119-126.
- HIRST, W. (1982). The amnesic syndrome: Descriptions and explanations. *Psychological Bulletin*, **91**, 435-460.
- HOFFMAN, H. G. (1997). The role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 371-383.
- HOFFMAN, H. G., GARCIA-PALACIOS, A., THOMAS, A. K., & BOTELLA, C. (2001). *Virtual-reality monitoring: Phenomenal characteristics of real, virtual, and false memories*. Manuscript submitted for publication.
- JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.
- JOHNSON, M. K., FOLEY, M. A., SUENGAS, A. G., & RAYE, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, **117**, 371-376.
- JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- JOHNSON, M. K., NOLDE, S. F., & DE LEONARDIS, D. M. (1996). Emotional focus and source monitoring. *Journal of Memory & Language*, **35**, 135-156.
- JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review*, **88**, 67-85.
- JOHNSON, M. K., RAYE, C. L., FOLEY, M. A., & KIM, J. K. (1982). Pictures and images: Spatial and temporal information compared. *Bulletin of the Psychonomic Society*, **19**, 23-26.
- JOHNSON, M. K., & SUENGAS, A. G. (1989). Reality monitoring judgments of other people's memories. *Bulletin of the Psychonomic Society*, **27**, 107-110.
- KASSIN, S. M., & KIECHEL, K. L. (1996). The social psychology of false confessions: Compliance, internalization and confabulation. *Psychological Science*, **7**, 125-128.
- KWONG SEE, S. T., HOFFMAN, H. G., & WOOD, T. (in press). Perceptions of the older witness: Is the older eyewitness believable? *Psychology & Aging*.
- LINDSAY, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1077-1083.
- LINDSAY, D. S., & JOHNSON, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, **17**, 349-358.
- LOFTUS, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- LOFTUS, E. F., & HOFFMAN, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General*, **118**, 100-104.
- LOFTUS, E. F., MILLER, D. G., & BURNS, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 19-31.
- LOFTUS, G. R., & LOFTUS, E. F. (1988). *Essence of statistics* (2nd ed.). New York: Knopf.
- MCINTYRE, J. S., & CRAIK, F. I. M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology*, **41**, 175-192.
- NELSON, T. O., KRUGLANSKI, A. W., & JOST, J. T. (1998). Knowing thyself and others: Progress in metacognitive social psychology. In V. Y. Zyerby, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 69-89). Thousand Oaks, CA: Sage.
- ROEDIGER, H. L., III, MEADE, M. L., & BERGMAN, E. T. (in press). Social contagion of memory. *Psychonomic Bulletin & Review*.
- SCHACTER, D. L., HARBLUCK, J. L., & MCLACHLAN, D. R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning & Verbal Behavior*, **23**, 593-611.
- SCHNEIDER, D. M., & WATKINS, M. J. (1996). Response conformity in recognition testing. *Psychonomic Bulletin & Review*, **3**, 481-485.
- SMITH, V. L., & ELLSWORTH, P. C. (1987). The social psychology of eyewitness accuracy: Misleading questions and communicator expertise. *Journal of Applied Psychology*, **72**, 294-300.
- STRACK, F., & BLESS, H. (1994). Memory for non-occurrences: Metacognitive and presuppositional strategies. *Journal of Memory & Language*, **33**, 203-217.
- WELLS, G. L., & BRADFIELD, A. L. (1999). Distortions in eyewitness recollections: Can the postidentification-feedback effect be moderated? *Psychological Science*, **10**, 138-144.

(Manuscript received May 24, 1999;  
revision accepted for publication September 13, 2000.)