In Experiment 1 (N = 16), under conditions of high memory load (60 pictures and 50 paragraphs) and a 1-week retention interval, undergraduate subjects reported their memory for photographs of scenes (card recall and free-recall tasks). Subjects frequently reported memory for photographs that they had actually never seen, but had read about in a brief paragraph. In Experiment 2 (N = 40), the same pattern of results was obtained with immediate testing. Experiment 1 also demonstrated that the likelihood of subjects falsely attributing scene memory (based on reading) to actually having viewed a photograph was reduced when metacognitive awareness of imaging during reading was made salient. Awareness of image creation was induced by requiring subjects to rate the paragraphs with respect to imagery vividness. Although other measures of memory remained the same, subjects in the induced-imagery condition made 50% fewer confusion errors than subjects who read the paragraphs without imagery instructions. The results are discussed in the context of Johnson and Raye’s (1981) reality monitoring model.

We remember visual scenes that we have perceived and those that we have imagined. Imagination of a scene may occur during dreaming, listening to a story, and also quite commonly during reading. In the present research we studied the observer’s ability to determine the origin of a remembered scene. Subjects had to decide if a remembered visual scene was one they had seen in a photograph (perceived) or one they had read about (imagined). Johnson and Raye (1981) refer to this decision process as “reality monitoring.” Although individuals typically have little difficulty in successfully monitoring the source of their memories, errors or equivocations in the reality monitoring process sometimes occur. For example, an adult may recall a vivid childhood memory, only to discover that he or she had actually not been present when the event took place, but had heard it recounted by others on numerous occasions. Equivocations in attributing the source of a remembered event occasionally occur upon waking when one tries to determine if an event was dreamed about or actually had occurred. In addition to interesting anecdotal accounts of this phenomenon,

AMERICAN JOURNAL OF PSYCHOLOGY
© 1992 by the Board of Trustees of the University of Illinois
realify monitoring errors have been documented in the laboratory. Johnson, Raye, Foley, and Foley (1981) reported such errors when subjects were required to indicate if a remembered word had been presented or generated on the basis of partial letter information. Anderson (1984) reported similar types of errors when subjects were required to indicate whether they had made, watched, or imagined simple motor movements. Johnson, Kahan, and Raye (1984) reported reality monitoring errors when subjects were asked to indicate whether they or their partner had reported various types of stories (actual dreams, dreams they read, or dreams they made up). Johnson and Raye (1981) provide a framework to account for these errors.

According to the model, memories of imagined events differ from memories of external events along four dimensions: (a) amount of spatial and temporal contextual attributes, (b) number of sensory attributes, (c) number of specific details about the event, and (d) degree of awareness of the cognitive operations associated with the event. Generally speaking, perceived events are thought to be richer in the first three dimensions than are imagined events. Compared with imagined events, perceived events are thought to have relatively limited information regarding knowledge of cognitive operations, because the operations involved in perception are generally rapid and automatic. To the extent that cognitive operations are more salient and other information less detailed, subjects should be more likely to attribute the memory to an internal source (i.e., imagination). To the extent that metacognitive awareness is limited, and other information relatively detailed, subjects should be more likely to attribute the memory to an external source (i.e., perception).

Our purpose was to study reality monitoring as it applies to visual memory and reading. When reading a description of a scene, one can imagine the spatial layout and the perceptual characteristics of the scene. We were interested in the reality monitoring errors that might arise when a subject must discriminate memories of scenes based on such descriptions from those based on having viewed photographs. In Experiment 1 we report a procedure that results in monitoring errors of this kind. Experiment 2 was designed to determine if the error rate obtained with this procedure would be affected by varying the salience of metacognitive operations instituted during initial reading of the descriptive prose.

EXPERIMENT 1

Experiment 1 was designed to determine if under conditions of high memory load, we could induce reality monitoring errors using
REMEMBERING SCENES THAT WERE NEVER SEEN

relatively rich and complex stimuli. The stimuli were color photographs of scenes and verbal descriptions of scenes. Specifically, we sought to determine if subjects would remember having seen a color photograph of a scene they had read about, but had not viewed in pictorial form. The possibility that subjects would make such errors with any reliable frequency seemed sufficiently remote that we decided to test subjects after an extended retention period. As a starting point, we selected a 1-week delay.

Subjects were told a cover story that they would take part in two different experiments: a picture memory experiment and a reading comprehension experiment. In the first experiment, they studied a series of 60 pictures for about 5 s each. To validate the cover story, subjects were asked four recognition questions about the picture set. They were then introduced to the second "experiment," in which 50 question-and-paragraph pairs were presented on a monitor, one at a time. Subjects read each paragraph and answered the associated question; 10 of those paragraphs described 10 of the pictures they had seen in the picture experiment.

The following week, subjects returned with the expectation that they would be doing the same two experiments with stimuli presented at a faster rate. Instead, they were informed that in actuality their memory for the previous week's paragraphs would be tested in a recognition test. They were told that they might have noticed that a few of the paragraphs they had read described pictures that had been shown in the picture memory experiment. Given that the subject recognized a paragraph, he or she was asked to indicate if that paragraph had been associated with a picture that had been shown. After the recognition test, to obtain a free measure of the subjects' memory for the photographs, they were asked to think back to the picture memory experiment and describe as many of the photographs as they could.

The major questions were (a) whether subjects would correctly identify those paragraphs that actually had described pictures, and (b) whether subjects would, with any reliable frequency, report having seen photographs of scenes that had been described verbally but had not been presented in photographic form.

METHOD

Subjects

Seventy volunteers from the University of Delaware (9 women) were paid $6.00 for participating.
Stimuli

The stimuli were 110 visual scenes and 60 paragraphs each containing two to four sentences. The visual scenes were selected from the set of magazine photographs collected by Potter and Levy (1969) and depicted a wide variety of indoor and outdoor scenes. The 60 paragraphs were created by describing 30 of the 110 scenes. The remaining 30 scenes did not have associated paragraphs. They were selected by four judges who screened them to remove any pictures that might be confused with the scenes described in any of the 60 paragraphs.

Apparatus

Paragraphs were presented on the monitor of an Apple II-E Plus computer. Scenes in color were projected using a Kodak Carousel slide projector onto a 3.9 x 26-cm screen that fit over the monitor. This was done so that pictures and paragraphs would be presented in the same spatial location, thus removing a potential discriminating contextual cue regarding the subjects' viewing experience.

Procedure

Subjects were tested individually. They were told that they were going to take part in two experiments: a picture memory experiment and a reading comprehension experiment. In the "picture memory" experiment, subjects viewed 60 scenes for approximately 5 s each. They were instructed to remember as many scenes as possible because memory for the scenes would be tested. In accord with this cover story, after presentation, they were asked four questions about the pictures. For example, "Did you see a picture with a penguin in it?" Two questions yielded yes responses and two yielded no responses. The "reading comprehension experiment" was then described. Each subject was presented with 50 paragraphs (one at a time) on the computer monitor. Each paragraph was presented before a reading comprehension question. Subjects read each question and paragraph aloud and then answered the question. The experimenter entered the verbal response on the keyboard. An example of one question/paragraph pair is: "What is in the distance?" "The view was majestic. Rising behind the orchard and the meadow one could see the snow-covered peak. Actually much more than the peak was covered with snow, which made me think about the special equipment one would need to climb it." The subject's response would be "monosyllabic."

Forty of the paragraphs described pictures that had also been presented in the picture presentation phase of the experiment. Ten of the paragraphs did correspond to pictures that had been presented in the picture presentation phase. These were randomly chosen from the 60-item picture presentation set for each of the eight orders of paragraph presentation used in the experiment (2 subjects received each order). After the reading comprehension phase, subjects were scheduled to return the following week. They were told that they would take part in two similar experiments in which the stimuli would be presented at a faster rate.
When subjects returned the following week, they received an unanticipated recognition memory and source attribution test. They were presented with all 56 stimulus paragraphs (40 that did not correspond to any of the pictures, and 16 that did correspond to 10 of the pictures), one at a time, mixed with 16 foils (paragraphs they had not read before or seen in pictorial form). This yielded 60 paragraphs. Each was presented individually on the computer screen. Subjects were asked to indicate whether or not they had read the paragraph before, and to provide a confidence rating on a scale ranging from sure (3), pretty sure (2), to not sure (1). If the answer was no, they were told to think back to the picture memory experiment and indicate (yes or no) if they had seen a photograph of the scene described in the paragraph and to indicate their confidence (same scale).

It is important to recall that 18 of the 50 paragraphs actually did describe pictures that the subjects had seen before. In the instruction for the recognition test, subjects were given an example of a picture and a corresponding paragraph and were told that they may have noticed during reading comprehension that occasionally a paragraph corresponded to one of the previously viewed pictures in the "picture memory experiment."

After the recognition test, subjects were instructed to think back to the photographs in the picture memory experiment and to write a description of each photograph they could recall. To avoid any ambiguity, they were told to be careful not to confuse the photographs with scenes they had only read about. These were explicitly told that we were interested in their recall of the color photographs.

RESULTS AND DISCUSSION

During the presentation phase, subjects apparently had little difficulty understanding the paragraphs; only 2% of the answers to paragraph questions were unclear or wrong.

Recognition

After a 1-week delay, subjects correctly recognized 79% of the paragraphs. The false-alarm rate for paragraph recognition was 6%. Subjects were very good at identifying those pictures and paragraphs that had actually corresponded. Of the 10 randomly selected paragraphs falling into this category, subjects correctly reported that there had been an associated picture 71% of the time with a mean confidence rating of 2.6 (SD = .37). The most interesting outcome of the experiment, however, was that upon rereading the 40 paragraphs that had not been associated with a picture, 38% of the time subjects reported that they had seen a color photograph depicting the described scene. The mean confidence rating for the nonexistent scenes was 1.9 (SD = .37).
The difference in confidence between the two conditions was significant, \( t(15) = 5.02, p < .001 \). This is not surprising given that memory for photographs that had been visually perceived would be expected to contain more detail than memory for the imagined photographs. What is interesting is that subjects' confidence ratings for the non-imagined pictures were often quite high; the mean rating of 1.9 corresponds to a rating of "pretty sure." In sum, the notable aspects of these results are (a) the frequency of the errors, (b) the fact that all 16 subjects made them, and (c) the level of confidence subjects reported for the non-imagined photographs.

A similar phenomenon was reported by Dellet and Wilcox (1968) in a recognition memory experiment in which subjects were presented with a mixed sequence of photographs and descriptions. During the recognition test, subjects had to indicate whether the test item was in the same form or had been presented in the opposite form. Subjects sometimes misidentified descriptions as having been presented in photographic form. They rarely made the opposite type of error. This asymmetrical tendency could not be attributed to response bias (i.e., a tendency to say "picture" when unsure) because false alarms were almost equally designated as pictures and as descriptions.

**Free recall**

To determine if the types of errors just described occur only in source attribution tasks, after the recognition test, subjects took part in a free-recall task. The purpose was to obtain a free, unconstrained measure of memory that might be less susceptible to response bias than the yes/no task. The subjects were instructed to think back to the photographs that had been projected on the screen in the "picture memory experiment" and to recall as many as possible. They were told to be careful not to confuse the stimuli from the two experiments—that we were interested only in memory for the color photographs in the picture memory experiment. Subjects were then given as much time as they required to recall the photographs. The average number of scenes subjects recalled was 13.2 (SD = 4.5). Of these self-generated descriptions, only 54% were scenes that had actually been presented, 29% were scenes that had not been presented in photographic form but had been read about, and 17% could not be unambiguously attributed to any single scene or paragraph by the experimenters because the description was (a) too vague, (b) contained components from several scenes, or (c) appeared to be unrelated to any picture.

The results show that after the recognition-source-attribution task, subjects freely described their memory for photographs that had
In Experiment 2 we sought to determine if the subjects’ apparent failures in reality monitoring would be affected by one of the dimensions described in the Johnson and Raye (1981) model. According to the model, one of the cues used to distinguish memories based on external stimulation from those based on imagination is the record of cognitive operations. Therefore, an image that is incidentally created is more likely to be confused with a past perception than an image that is intentionally created, because memory for the intentionally created image contains information pertaining to the conscious act of imagining.

For example, Durso and Johnson (1980) presented subjects with a series of written nouns and outline objects. They asked three types of orienting questions during acquisition: verbal (e.g., naming); imaginal (e.g., “Create an image. How ‘good’ is the image?”); and referential (e.g., “How large is the real world object?”). Memory was tested later with a list of nouns presented auditorily. Subjects had to indicate if each auditorily presented object had been included in the visual list and if it had been presented as a picture or as a word. Results showed that both verbal and imaginal questions led to an equal number of both types of errors (calling a picture a word, and calling a word a picture). In the referential condition, however, subjects were more likely to call a picture a word than vice versa. Johnson and Raye (1981) account for this result by assuming that in answering the referential question, the subject incidentally created an image, resulting in a visuospatial memory without a metacognitive record that imaging had taken place. In the explicit imagery condition, however, not only is the visuospatial representation remembered, but also awareness of the act of imagining—a record of the particular associations activated during that process. This account was further supported by the observation that reality monitoring errors are less likely in a difficult imagery comparison task than in a relatively easy one, presumably because the more difficult task is associated with
greater awareness of cognitive operations (Finke, Johnson, & Shyi, 1988).

According to this argument, if confusion errors obtained in Experiment 1 are due to failures in reality monitoring, we should observe a reduction in confusion errors when subjects are explicitly required to create images of the scenes depicted in the reading comprehension paragraphs. This prediction rests on the assumption that relative to the metacognitive dimension, information on the other reality monitoring dimensions will remain roughly equal.

On the other hand, if confusion errors are dependent on how large and vivid the set of visual memories is, then we should observe an increase in the number of confusions when subjects are required to create images of each of the 50 paragraphs. This prediction assumes that although subjects might create an image of a scene when answering a question about it, sometimes they might not (cf. Paivio, 1971). Requiring the subject to create and rate an image of each description would increase the number of stored visuospatial representations and perhaps their quality and durability as well, thus increasing the likelihood of false picture memories.

Experiment 2 was designed to determine if imagery instructions would affect the frequency of confusion errors, and if so whether this would lead to an increase or a decrease in the probability that subjects would falsely attribute their scene memory to having viewed a color photograph. A secondary purpose was to determine if we could replicate the basic finding (which involved a 1-week delay) within a single experimental session.

METHOD

Subjects
Forty undergraduate volunteers (men and women) were paid for participating. There were 20 subjects in each condition.

Stimuli and apparatus
These were the same as in Experiment 1.

Procedure
There were two conditions: incidental imagery and intentional imagery. The procedure for subjects in the incidental imagery condition was the same as in Experiment 1, except that the acquisition and test phases were presented in the same experimental session. As in Experiment 1, during the reading comprehension phase, subjects read each question and paragraph out loud and then answered the question. This was followed by the paragraph recognition test and source attribution test (the recall task was omitted.
RESULTS AND DISCUSSION

As in Experiment 1, subjects in the incidental imagery condition had little difficulty understanding the paragraphs; only 6% of their responses to the associated questions were unclear or wrong. As expected, reducing the retention interval from 1 week (Exp. 1) to one session (Exp. 2) resulted in better paragraph recognition. The proportion of paragraphs recognized and the false-alarm rate for paragraph recognition in the incidental imagery condition were .92 and .01, respectively, and in the intentional imagery condition they were .93 and .02, respectively.

It is clear that imagery instructions had no effect on paragraph recognition. The proportions of hits and false alarms were virtually identical. Similarly, subjects in both groups were accurate and confident in identifying those 10 paragraphs that actually had been associated with previously viewed pictures. The proportion of correct picture reports in each condition is presented in Table 1. No difference between the groups was obtained, \( t(18) = 1.03 \). Also in Table 1, are the proportions of false picture reports for both conditions. Consistent

<table>
<thead>
<tr>
<th>Response type</th>
<th>Incidental</th>
<th>Intentional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>Conf</td>
</tr>
<tr>
<td>Correct</td>
<td>.97</td>
<td>2.7</td>
</tr>
<tr>
<td>False</td>
<td>.21</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 1. Proportion (P) of correct picture reports and false picture reports in the incidental and intentional imagery conditions and the mean confidence (Conf) rating associated with each.
with the reality monitoring model, this was the only dependent measure that did show a marked difference between the intentional and incidental imagery groups. Subjects yielded twice as many false picture reports (cases in which the subject reported a picture when no picture had been presented) in the incidental imagery condition as in the intentional imagery condition, \( t(38) = 2.81, p < .01 \). According to the model, when memory for an event includes metacognitive information about imagery that took place when the event was initially encountered, subjects should be less likely to attribute that memory to an external source. Clearly, subjects who had rated the vividness of the images associated with the paragraphs made fewer errors in reality monitoring that subjects who had read the paragraphs and answered a question. The latter group of subjects was more likely to erroneously remember a scene as having been presented in pictorial form.

An alternative interpretation of these results is that subjects had suppressed false picture reports in the intentional imagery condition because of response bias. That is, because subjects were aware of creating many images, they might have been less likely than subjects in the incidental imagery group to report "I saw a picture" when they were unsure, if a criterion shift is responsible for the difference. If false picture reports, then one would expect to see fewer correct picture reports in the intentional imagery group than in the incidental group, yet the groups did not differ on this measure. Both groups of subjects were accurate and confident in detecting those paragraphs that had been associated with photographs.

The high accuracy in picture recognition, however, raises the possibility of a ceiling effect, so that the lack of a difference here is suggestive but not conclusive. However, a second measure, confidence rating, is also relevant to the issue. If response bias in the intentional imagery condition causes subjects to be more conservative in saying "I saw a picture," then the mean confidence rating should be higher in the intentional imagery condition than in the incidental imagery condition. As shown in Table 1, this was not the case for the correct picture reports. Subjects in both groups were very confident of these responses. The frequency use of the highest rating, however, again raises the issue of a ceiling effect, but for the false picture reports, there clearly was no ceiling effect (the mean response was at the center of the scale) and the groups did not differ. An analysis of variance (ANOVA) (Picture Report [hits vs. confusions] \( \times \) Group) conducted on the confidence ratings showed that \( \chi \) in Experiment 1, subjects were less confident when they made a false picture report than when they correctly reported a picture that actually had been
presented, \( F(1, 55) = 91.73, MS_E = 12.29, p < .0001 \); however, that confidence did not differ with imagery condition, \( F < 1 \), nor was there an interaction, \( F(1, 55) = 1.90 \).

**GENERAL DISCUSSION**

Reality monitoring refers to the processes that allow the individual to determine if a remembered event is based upon external experience or imagination (Johnson & Raye, 1981). In two experiments, we tested a method for inducing errors in reality monitoring. Under conditions of high memory load, subjects frequently and confidently reported having seen photographs of scenes when no photograph depicting that scene had been presented. Reading a prose description of a particular scene sometimes resulted in a reported “picture memory” for the scene. The phenomenon occurred both after a 1-week retention interval (Exp. 1) and within a 1-hr experimental session (Exp. 2).

The primary question is whether the subjects’ false picture reports actually indicate a failure in reality monitoring (i.e., that the subjects actually believed they had seen the nonexistent pictures), or whether they were the result of response bias inherent in the task itself. Several aspects of the results, as well as a compelling, unexpected consequence of conducting these experiments on the participants themselves (discussed below), strongly suggest that these are indeed reality monitoring errors as discussed by Johnson and Raye (1981).

The main concern is that the source-attribute test itself, by requiring subjects to indicate if a picture had been associated with each of the recognized paragraphs, might have biased the subjects to report yes. Two aspects of the test procedure were designed to minimize any such bias. Ten paragraphs were included that in actuality did correspond to pictures, and at test in both experiments, subjects were told that they may have noticed that occasionally one of the paragraphs in the reading comprehension test had described a photograph that they had seen in the picture memory experiment. This latter assertion described the situation accurately in that picture-paragraph correspondence occurred for only 10 of the 50 initial paragraphs. This was deliberately stressed to help reduce a bias for reporting many pictures. The very large number of reported pictures (pictures were reported for more than half the paragraphs) coupled with the surprisingly high confidence rating associated with the false picture reports argues against a simple response bias explanation of the report.
Further evidence against a response bias explanation is the preponderance of false picture reports during the subjects' free recall of the photographs presented in the picture presentation phase (Exp. 1). Subjects in this case were instructed to think back to the photographs and describe as many as they could remember. They were warned not to confuse pictures with paragraphs. The recall results, which were open-ended and subject generated, suggest true confusion errors. All subjects described pictures they had never seen and, as reported earlier, a few proceeded to draw their recollections of the nonexistent photographs.

In Experiment 2, we found that the number of false picture reports was affected by imagery instructions in the direction predicted by the reality monitoring model. Explicit imagery, which we argue is more likely to yield a metacognitive record of imaging than the implicit imagery that may occur during reading, led to a reduction in the number of false picture reports. Whereas all other measures of memory remained constant, the incidental imagery group made twice as many false picture reports as the intentional imagery group. A response bias account of the reduction in error rate was not supported. No difference in confidence was obtained between the false picture reports in the two conditions, as would be expected if subjects in the two conditions used different rejection criteria. Nor was there a difference in the hit rate for correctly identifying those paragraphs that had been associated with pictures. Subjects simply made fewer false picture reports.

Additional support that the false picture reports are reality monitoring errors was obtained in an unusual way. In preparing the stimuli for the experiment, the authors unintentionally experienced conditions very similar to those of the experiment. We viewed a subset of pictures one week, helped the assistants construct paragraphs for the pictures on various occasions in the ensuing weeks (both with and without the pictures being present), and then recalled the pictures a few days later. As a result of these activities we both remembered many photographs. Of interest, however, was the fact that we each remembered a sunny, brightly colored photograph of a father and son at the zoo, in which the father held up his young son to look at an elephant. The son was wearing a red and white striped shirt. We were each confident about our recall of the picture, but in recounting our descriptions found great discrepancies regarding the orientation of the objects, most notably the elephant. We consulted the photograph to determine who was correct and were both amazed to find a completely unfamiliar picture depicting the father, son, and elephant at the zoo on a drab winter's day.
What had happened was that neither of us had seen that picture. The scene had been described to one of us by a research assistant who was having a problem writing an associated paragraph. In discussing the problem and creating an acceptable paragraph, we had apparently made implicit images of the description that we later believed to be based on having viewed the photograph. Neither the weather nor the clothing had been referred to in the description. The bright, sunny day is a detail that most people would guess about a picture of a father and son at the zoo. The red and white striped shirt was present in another picture that we actually had seen. Furthermore, the knowledge that the images we recalled were based on verbal descriptions of the scene did not change the subjective sense that they were recollections of a photograph.

In conclusion, these experiments provide additional support for the reality monitoring model of Johnson and Raye (1981). They show that reality monitoring errors can be reliably induced for rather complex and detailed visual information. Previous research has shown that subjects’ memory for pictorial details can be altered by presenting additional information to them verbally (e.g., Perzik, 1977; Schoolder, Gerhard, & Loftus, 1986). The current experiments show that subjects sometimes remember having seen a picture they had actually never seen but had only read about. This demonstrates that reality monitoring can be studied using relatively rich and complex stimuli. The approach taken in these experiments may be useful in future research into reality monitoring, imagery during reading, and metacognition.

Notes
This research was presented at the 27th annual meeting of the Psychonomic Society, New Orleans, LA, November 1986. Preparation of this article was supported in part by National Science Foundation Grant BNS-8519417, Helene Intraub, and Army Research Office Contract DAAI.05-86-A-0080 to James E. Hoffman. We thank Alan Scott and Karen Burns for their valuable assistance in conducting the experiments. Correspondence concerning this article should be addressed to Helene Intraub or James E. Hoffman, Department of Psychology, University of Delaware, Newark, DE 19716. Receives for publication November 8, 1990; revised received March 18, 1991.

1. As an exploratory variable, paragraphs were presented in two slightly different versions (explicit and inferential) that occurred equally often in the experiment. Because this variable had no effect on any of our reality monitoring measures, the data from these two versions were combined. An example of the explicit version is “What is in the distance?”/”The view was majestic. Rising behind the orchard and the meadow one could see the
snow covered peak of the mountain. Actually much more than the peak was covered with snow, which made me think about the special equipment one would need to climb it." The incidental version was identical except that the phrase "snow covered peak of the mountain" was replaced with "snow covered peak."

2. See note 1.

References
Andersen, R. E. (1994). "Did I do it or did I only imagine doing it?" Journal of Experimental Psychology: General, 113, 594–615.