

Studying related pictures can reduce accuracy, but increase confidence, in a modified recognition test

C. C. CHANDLER

Washington State University, Pullman, Washington

In 14 experiments, a dissociation occurred between subjects' accuracy on a forced-choice recognition test and their confidence in their choice. Nature pictures (e.g., Lake A) were shown. Later, the subjects were asked to choose the picture that they remembered, given the target (Lake A) and a novel picture (Lake C) as alternatives, and rated their confidence in their choice. When the subjects also studied a related picture (Lake B), their accuracy often decreased while their confidence increased. The dissociation cannot be explained by signal detection theories of recognition, which assume that strength determines both accuracy in a forced-choice test and confidence. Instead, familiarity with general themes may give people the illusion that they are accurately remembering details.

When people say that they remember an event, they are referring to a subjective feeling that the event occurred in their past (e.g., Tulving, 1983). Traditionally, cognitive psychologists have been interested in measuring the accuracy of memory, but they have largely neglected subjective confidence. It is generally acknowledged that subjects must have some degree of confidence in order to say that they studied an item. However, it is unusual for researchers to ask what variables influence subjective confidence and whether the same factors determine accuracy.

Understanding the relation between accuracy and confidence is important from a practical, as well as a theoretical, standpoint. It has been suggested that people use their subjective confidence to gauge the accuracy of their recollection (e.g., Jacoby, Kelley, & Dywan, 1989; Nelson & Narens, 1990). Degree of confidence also determines our consciously directed actions (e.g., Nelson & Narens, 1990). For example, you would greet someone enthusiastically only if you are confident that you remember that person. The perceived accuracy of the recollection, gauged by subjective confidence, guides our actions.

Furthermore, people who seem confident of their recollections are also judged to be accurate. In judging eyewitness testimony, it has been shown that jurors are more likely to believe the witnesses whom they perceive as being confident (R. C. L. Lindsay, Wells, & Rumpel, 1981; Wells, Ferguson, & R. C. L. Lindsay, 1981). Also, jurors are more likely to believe witnesses who rated themselves as being confident (Wells et al., 1981). Jurors'

inferences regarding a witness's accuracy are largely determined by the confidence of the witness. When a witness is highly confident, jurors ignore situational factors that would decrease the witness's accuracy (R. C. L. Lindsay et al., 1981). Thus, from a practical standpoint, it is important to determine whether variables affect confidence and accuracy differently.

From a theoretical standpoint, subjective confidence has been regarded as a measure of memory strength. The signal detection theory of recognition (e.g., Snodgrass, Levy-Berger, & Haydon, 1985, Appendix B) typifies this view. The theory assumes that both studied and nonstudied items vary on a dimension referred to as "strength." In an old/new recognition test, an item is judged to be "old" (or "studied") if its strength exceeds a criterion set by the subject. It is assumed that the subject sets the criterion at a particular strength. When confidence ratings are taken, it is assumed that each number represents a criterion that corresponds to a particular strength.

According to signal detection theory, the difference in the strengths of studied and novel items determines one's ability to discriminate between them. The ability to discriminate between the studied and novel items can be measured by the percentage of correct responses in a forced-choice recognition test (e.g., Green & Swets, 1974, chap. 2). Likewise, confidence in one's choice should reflect the difficulty in choosing between items on the test. Any variable that increases the overlap in the strengths of studied and novel items should (1) reduce accuracy in the forced-choice recognition test, and (2) reduce confidence in one's choice. Contrary to this prediction from signal detection theory, the present article reports a variable that can reduce accuracy while increasing confidence. The dissociation shows that confidence ratings are not solely determined by memory strength. By studying what determines confidence, as well as what determines ac-

Portions of this study were reported at the 1992 meeting of the American Psychological Society in San Diego, California. Address all correspondence to C. C. Chandler, Department of Psychology, Washington State University, Pullman, WA 99164-4820 (e-mail: chandler@wsuvm1.csc.wsu.edu).

curacy, we will achieve a better understanding of subjective remembering.

What factors might influence accuracy and confidence in opposite ways? Dissociations between accuracy and other measures of metamemory may provide some clues. Metamemory is defined as a person's knowledge or beliefs about the accuracy of his/her own memory. Measures of metamemory include feeling-of-knowing judgments (FOK) and tip-of-the-tongue ratings (TOT), as well as confidence ratings. Some studies have reported dissociations between accuracy and measures of metamemory (e.g., FOK judgments). For example, Nelson, McSpadden, Fromme, and Marlatt (1986) measured recall of general knowledge questions (e.g., "What is the capital of Chile?"). When the subjects could not recall the answer, they predicted whether they could recognize the answer (an FOK judgment). Alcohol intoxication decreased recall but did not affect FOK judgments. In contrast, Nelson et al. (1990) reported that exposure to extreme altitudes (Mount Everest) did not affect subjects' recall, but did lower FOK judgments. Thus, physiological states dissociate accuracy and FOK judgments.

Koriat and Lieblich (1977) suggested that some measures of metamemory can be affected by cognitive factors that are independent of memory for a target event. FOK judgments and TOT states, for example, are influenced by the familiarity of retrieval cues (see, e.g., Metcalfe, Schwartz, & Joaquim, 1993; Reder & Ritter, 1992). For example, Metcalfe et al. presented word pairs (A-B) as targets and supplied the first word in the pair (A) as a cue for recall. The familiarity of the cue (A) was manipulated by presenting that cue once (C-D, A-B condition) or twice (A-D, A-B condition) during the study phase. Recall was equivalent for the two conditions, but greater cue familiarity was associated with higher FOK judgments (Experiments 1, 2, and 3) and more TOT states (Experiment 4). Increasing the familiarity of a cue increases the feeling that subjects know, and can remember, the target.

Reder and Ritter (1992) found a similar dissociation between accuracy and FOK judgments when they asked their subjects to perform mathematical calculations ($23 \times 34 = ?$). Similar problems were presented in the test phase, and the subjects were asked to quickly indicate whether they could retrieve the answer from memory, or whether they would have to calculate the answer (an FOK judgment). Then, a short period of time was allowed to answer each question. The subjects were more likely to say that they could retrieve the answer if some part of the problem was familiar ($23 + 34 = ?$). The more frequently the parts had been presented, the more "retrieve" responses were given, even though accuracy never increased.

A greater familiarity with the cue increases both TOT states (Metcalfe et al., 1993) and FOK judgments (Reder & Ritter, 1992), without increasing accuracy. Would a greater familiarity with the cue also increase confidence

without increasing accuracy? Like FOK and TOT judgments, confidence is a measure of metamemory and may rely on factors that generally influence measures of metamemory. However, FOK and TOT judgments *precede* a response, whereas confidence ratings *follow* a response, so confidence ratings may be based solely on details that have been retrieved about a specific target. Therefore, one should not assume that prospective tasks (FOK and TOT) and a retrospective task (rating confidence in one's recollection) measure the same thing. Indeed, Shimamura and Squire (1986, 1988) have shown a dissociation between confidence and FOK judgments. They presented sentences (e.g., "Sarah lost her keys") to people with Korsakoff's amnesia and to an alcoholic control group. Cued recall was equated for the groups by testing the alcoholics after a longer retention interval. The groups were equally confident in their correct answers and equally nonconfident in their incorrect answers (Shimamura & Squire, 1988). However, FOK judgments predicted recognition better for the alcoholics than for the subjects with Korsakoff's syndrome (Shimamura & Squire, 1986). Thus, the subjects with Korsakoff's syndrome were impaired on the prospective measure of memory (FOK judgments) without being impaired on a retrospective measure of memory (confidence in recall).

Shimamura and Squire's (1986, 1988) work shows that prospective and retrospective measures of metamemory are not necessarily influenced by the same factors. However, the present article provides evidence that confidence in one's recollection is influenced by the familiarity of the retrieval cue. Increasing the familiarity of a nature scene (e.g., a lake scene) can decrease accuracy in a forced-choice recognition test while increasing confidence in one's choice. This conclusion is based on accuracy and confidence data collected in 14 studies that employed an interference paradigm (Chandler, 1989, 1991).

My previous work focused on accuracy in order to determine what causes interference in a forced-choice recognition test (see Chandler, 1989, 1991). Confidence ratings were not reported because they were not relevant to the purpose of the articles. However, the data are relevant to the present purpose—namely, to show a dissociation between accuracy and confidence.

In all of the studies, two lists of scenic nature pictures were shown: a list of targets and a list of nontargets. Each target in the control condition was unrelated to the nontargets; each target in the experimental condition (e.g., Lake A) was related to one nontarget picture (e.g., Lake B). On the modified recognition test, the target was paired with a novel picture (e.g., Lake C), and the subjects were asked to choose the picture that had been shown. Upon making a choice, the subjects rated their confidence on a 6-point scale from *guessing* to *certain*. Across the 14 studies, seeing a related picture never increased the subjects' accuracy and sometimes impaired performance in the recognition test, yet seeing a related picture increased confidence.

METHOD

The 14 studies (described in Chandler, 1989, 1991) had similar methodologies. A typical experiment will be described, and then departures will be noted. (The reasoning behind the differences in methodology is discussed in Chandler, 1989, 1991.)

Materials

In all 14 experiments, the materials were nature scenes taken from Sierra Club calendars. Each scene was distinct from the others in subject matter; there was one scene of a lake thawing, one scene of sand dunes, and so forth. The materials for the experiments were made by cutting each scene into three pictures of equal size, and by mounting each picture on a 5 × 7 card (see Figure 1 for an example). For a given subject, one picture served as the target, another served as the nontarget, and the remaining picture served as the novel picture on the recognition test (see Figure 1).

Procedure

Two sets of pictures were shown—a set of target pictures and a set of nontarget pictures. Following a retention interval, a forced-choice recognition test was administered.

The target pictures (48 total) were shown at a 10-sec rate. As an orienting task, the subjects were asked to rate each picture on a 5-point scale (e.g., for complexity). The subjects also rated (e.g., for liking) a set of 48 nontargets that was shown at a fixed rate (e.g., 6 sec). Targets in the control condition were not related to any of the nontargets; however, each target in the experimental condition was related to one of the nontargets (e.g., both pictures were scenes of a lake). The list of nontargets was shown immediately

before the targets (proactive design) or immediately after the targets (retroactive design).

The two sets of pictures were followed by a retention interval (e.g., 15 min), and then by a forced-choice recognition test. The test paired each target with its *novel* counterpart (e.g., Lake A vs. Lake C), and the subjects were given an unlimited amount of time to select the picture that they had seen in the experiment. After choosing one of the alternatives, the subjects rated their confidence that they had made the correct choice. Confidence ratings ranged from 1 (*guessing*) to 6 (*certain*) on a Likert scale.

Subjects and Design

In each study, there were 18 to 36 students who were recruited from an introductory psychology course. Every subject responded to an equal number of items in the control and experimental conditions. In the control condition the targets were not related to any of the nontargets; in the experimental condition each target was related to one of the nontarget pictures. The design was either a proactive design, in which the nontargets preceded the targets, or a retroactive design, in which the nontargets followed the targets.

Across subjects, the scenes appeared in the control and experimental conditions equally often. Also, a given picture filled the roles of target, nontarget, and novel distractor equally often in each condition.

Variations in Methodology Across Experiments

The 14 studies varied in methodology in order to influence the amount of interference found in the forced-choice test (see Chandler, 1989, 1991). For easy reference, each study is identified by a different letter (A-N); procedure and design are summarized in

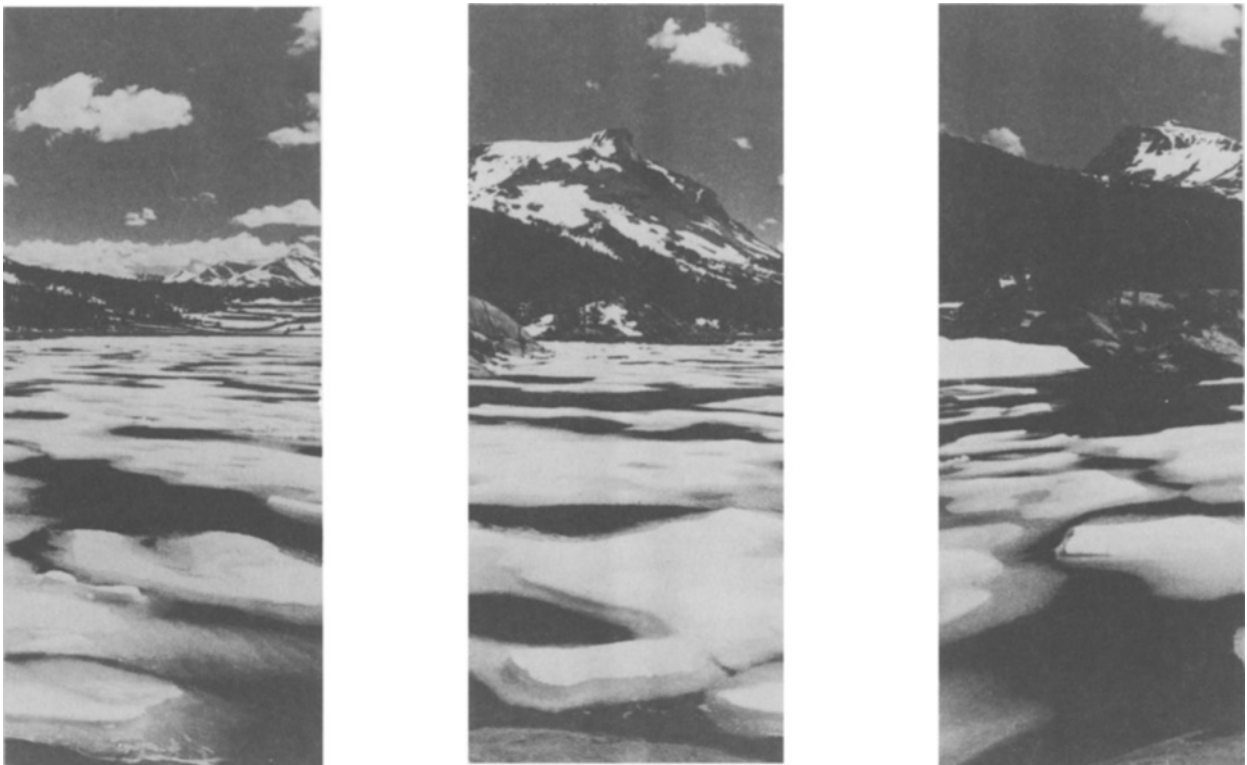


Figure 1. For a given subject, one picture served as the target, another served as the nontarget, and the remaining picture served as the novel distractor on the modified recognition test.

Table 1
Percentage of Correct Choices and Confidence Ratings in a Forced-Choice Recognition Test
as a Function of Methodology

Study	Design	Targets Presented/Rate*	Nontargets Presented/Rate*	Retention Interval	Percentage Correctly Recognized		Median Confidence Ratings		Sign Test <i>p</i> <
					Control	Experimental	Control	Experimental	
A	Retroactive 1989, Exp. 1	48/10 Memorability	48/6 Reminds?	10 min	80†	73	4.34	4.20	.419
B	Retroactive 1989, Exp. 2	48/7.5 Liking	48/6 Reminds?	10 min	79†	72	3.34	3.79	.271
C	Retroactive 1991, Exp. 1	48/10 Complexity	48/6 Liking	15 min	85†	74	3.40	3.54	.500
D	Retroactive 1991, Exp. 1	48/10 Complexity	48/10 Liking	15 min	81†	74	4.17	4.00	.500
E	Retroactive 1991, Exp. 5	48/10 Liking	48/10 Complexity	15 min	82†	74	3.92	4.25‡	.048
F	Retroactive 1989, Exp. 3	48/4 Memorability	48/6 Reminds?	10 min	65	64	2.32	2.83‡	.002
G	Retroactive 1989, Exp. 4	105/8 Memorability	105/6 Reminds?	10 min	68	70	3.42	3.86‡	.006
H	Retroactive 1991, Exp. 2	48/10 Complexity	48/6 Liking	48 h	80	81	3.22	3.73	.119
I	Retroactive 1991, Exp. 2	48/10 Complexity	48/10 Liking	48 h	80	78	3.64	4.21‡	.001
J	Proactive 1991, Exp. 3	48/10 Liking	48/6 Complexity	15 min	84	82	4.50	4.75	.119
K	Proactive 1991, Exp. 3	48/10 Liking	48/10 Complexity	15 min	84	84	4.83	5.04	.500
L	Proactive 1991, Exp. 5	48/10 Complexity	48/10 Liking	15 min	83	84	4.00	4.50	.159
M	Proactive 1991, Exp. 6	48/10 Liking	48/6 Complexity	48 h	81	79	3.25	3.77§	.079
N	Proactive 1991, Exp. 6	48/10 Liking	48/10 Complexity	48 h	80	79	3.63	4.00	.119

*Rate is given in seconds. †Significantly more accurate performance in the control condition than in the experimental condition ($p < .05$). ‡Significantly higher confidence ratings in the experimental condition than in the control condition ($p < .05$). §Marginally significant effect ($.05 < p < .10$).

Table 1. The table lists the type of design (retroactive or proactive), the rating task performed on the targets as well as their rate of presentation, the task performed on the nontargets and their rate of presentation, and the retention interval that separated the end of the second list and the test.

RESULTS

Seeing Related Pictures Can Decrease Accuracy but Increase Confidence

Analysis for individual studies. Table 1 shows the percentage of correct responses by condition (control or experimental) for each of the 14 studies (from Chandler, 1989, 1991). In five studies, performance was significantly more accurate in the control condition than in the experimental condition (A–E). In nine studies, performance was equally accurate in both the control and experimental conditions. The critical question is whether studying a related picture affects confidence in the same way.

Table 1 shows the median confidence ratings in the control and experimental conditions for each study. Each subject's median confidence rating was recorded for each condition, then an overall median was calculated for each condition. Seeing a related picture significantly *increased*

confidence in four studies and had a marginal effect in another study. In nine studies, confidence did not differ significantly for the control and experimental conditions.

In summary, seeing a related picture sometimes decreased a subject's accuracy in the forced-choice test. However, seeing the related picture never decreased confidence in a subject's choice; instead, it often increased confidence.

Meta-analysis. In a meta-analysis, I combined the data from studies that had the same design (retroactive or proactive) and retention interval (10–15 min or 48 h). The purpose of the meta-analysis was to increase power and test the generality of the effects. For each subject, I recorded whether accuracy was higher in the control condition (coded as positive), the experimental condition (coded as negative), or neither (coded as a tie). The tallies were entered into a sign test (see Lehmann, 1975, pp. 120–123). With large samples, the binomial distribution is approximated by the standard normal curve, so *z* scores are shown for each design and retention interval in Table 2.

A significant retroactive interference effect was found at the retention interval of 10–15 min. In other words, accuracy was higher in the control condition than it was

Table 2
Sign Test (z) for the Effect of Control/Experimental Condition
on Accuracy and Confidence Ratings

Design	Retention Interval	Effect on Accuracy ^a	Effect on Confidence Ratings ^b		
			All Choices	Correct Choices	Incorrect Choices
Retroactive	10-15 min	-3.75*	3.40*	2.91*	5.81*
High Accuracy (ABCDE)		-4.93*	1.46†	2.65*	7.52*
Low Accuracy (FG)		.29	3.90*	2.60*	3.18*
Retroactive	48 h	.00	3.17*	2.83*	4.00*
Proactive	10-15 min	-1.18	1.30†	.49	1.71*
Proactive	48 h	-0.67	2.00*	1.18	3.04*

^aNegative numbers indicate that accuracy was lower in the experimental condition. ^bPositive numbers indicate that confidence ratings were higher in the experimental condition. *A significant effect ($p < .05$). †A marginally significant effect ($.05 < p < .10$).

in the experimental condition. However, no retroactive interference was found after 48 h, and no proactive interference was found at either retention interval. The critical question is whether seeing the related picture had a similar effect on confidence ratings.

For each subject, I recorded whether confidence was higher in the control condition (coded as positive), in the experimental condition (coded as negative), or neither (coded as a tie), and entered the tallies into a sign test. Sign tests were chosen because they do not assume that people in different experiments use the rating scale in the same way. The positive z scores indicate greater confidence ratings in the experimental condition than in the control condition (see the fourth column of Table 2). In the retroactive design, the effect was significant at both retention intervals. In the proactive design, the effect was significant at the 48-h retention interval and was marginally significant at the retention interval of 10-15 min.

In summary, a dissociation was found between accuracy and confidence. In retroactive designs with retention intervals of 10-15 min, seeing related pictures decreased accuracy while increasing confidence. In all other cases,

seeing related pictures did not affect accuracy, but did increase confidence.

The Effect of Seeing Related Pictures on Confidence for Correct and Incorrect Choices

Further analyses showed that studying a related picture increased confidence for both correct and incorrect choices. Four mutually exclusive categories were created by crossing the two conditions (control, experimental) with accuracy of the decision on the forced-choice recognition test (correct choice, incorrect choice). A subject's datum was excluded from consideration if no responses occurred in one of the four categories; 0-2 subjects were excluded per experiment.

Analysis for all studies combined. Each subject's confidence ratings were sorted into the four categories (correct-control, correct-experimental, incorrect-control, and incorrect-experimental). Within each category, the percentage of decisions given each confidence rating (1-6) was recorded for each subject. These percentages were then used to calculate overall percentages for the 14 studies combined, which are shown in Figure 2.

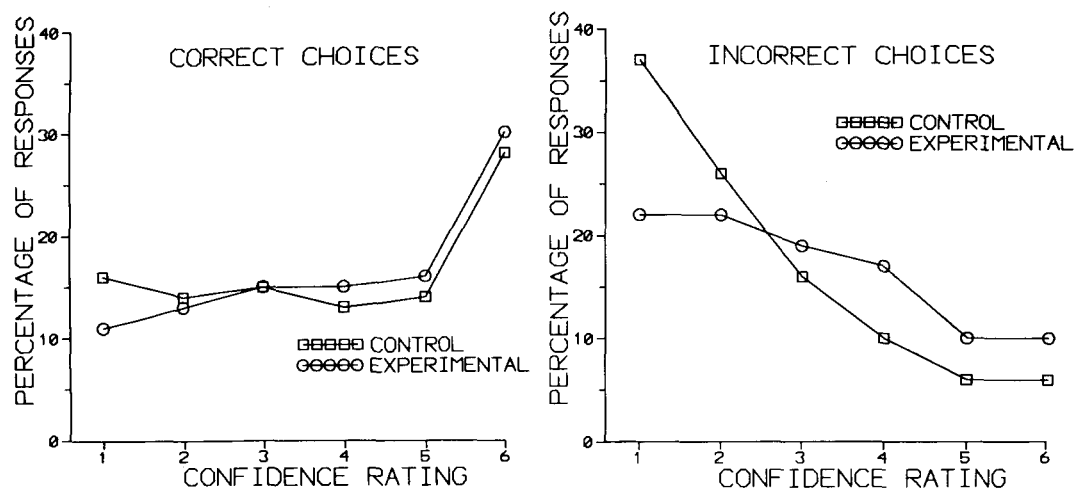


Figure 2. The distribution of confidence ratings for correct choices (left panel) and incorrect choices (right panel). Higher confidence ratings were given in the experimental condition than in the control condition.

Figure 2 shows that the subjects were more confident of their correct choices (left-hand panel) than they were of their incorrect choices (right-hand panel). Of the 300 subjects, 95% showed this pattern (sign test, $z = 15.65$, $p < .001$). The figure also shows that confidence ratings were higher in the experimental condition than they were in the control condition.

For correct choices, confidence tended to be higher in the experimental condition than in the control condition. In the experimental condition, the subjects gave a somewhat higher percentage of *certain* responses (a rating of 5 or 6) and a somewhat lower percentage of *guessing* responses (a rating of 1). Although the effect was small, it was statistically significant. Of the 300 subjects, 181 had higher median confidence ratings in the experimental condition, 118 had higher ratings in the control condition, and 7 were tied (sign test, $z = 3.87$, $p < .001$).

The effect of condition on confidence ratings was more dramatic for incorrect choices. Compared with the control condition, the experimental condition elicited a greater percentage of *certain* responses (a rating of 5 or 6), and a lower percentage of *guessing* responses (a rating of 1 or 2). Of the 300 subjects, 208 had higher median confidence ratings in the experimental condition, 74 had higher ratings in the control condition, and 18 were tied (sign test, $z = 7.68$, $p < .001$).

Analysis for each study. For each experiment, an overall median confidence rating was calculated for each of the four categories (correct-control, correct-experimental, incorrect-control, and incorrect-experimental). The medians are displayed in Table 3.

In 12 of the 14 studies, confidence in correct choices was numerically higher in the experimental condition than in the control condition. Sign tests showed that the effect was significant in 3 studies (see Table 3). Because the effect was small, there was not enough power to detect it in every experiment. In 13 of the 14 studies, confidence in incorrect choices was numerically higher in the ex-

perimental condition than in the control condition. Sign tests showed that the effect was statistically significant in 9 studies.

Meta-analysis. In a meta-analysis, I combined the data from studies that had the same design (retroactive or proactive) and retention interval (10–15 min or 48 h). Table 2 shows the effects of condition (control, experimental) on confidence ratings for correct choices (fifth column) and for incorrect choices (sixth column). In most cases, confidence ratings were higher in the experimental condition than in the control condition; the proactive designs provided the only exception. In the proactive studies, confidence ratings in correct choices were equal for the control and experimental conditions.

DISCUSSION

The popular signal detection theory of recognition assumes that discrimination measures (e.g., accuracy in a forced-choice test) and confidence ratings both measure memory strength. According to the theory, conditions that reduce accuracy in a forced-choice test should also reduce confidence ratings. The theory makes this prediction, because accuracy and confidence both reflect the difference in strength between studied items and nonstudied items. Any condition that decreases the difference in strength should reduce both accuracy and confidence ratings.

Contrary to these predictions, the present studies showed a dissociation between accuracy and confidence in a forced-choice recognition test (e.g., "Did you see Lake A or Lake C?"). In an immediate test, accuracy was reduced when the target (Lake A) was followed by a related picture (Lake B), but this decrease in accuracy was accompanied by an increase in confidence. In other circumstances, studying a related picture did not affect accuracy, but it increased confidence. These dissociations between accuracy and confidence argue against the idea that accuracy and confidence can be conceptualized solely in terms of memory strength.

Furthermore, the traditional measures of criterion (such as beta) do not capture important aspects of confidence. In an old/new recognition test, criterion indicates the likelihood that an observation was selected from the studied distribution versus the novel distribution. It is assumed that the subject can set many criteria, assigning a confidence rating to each one. The experimenter then uses the confidence ratings to plot an MOC curve that summarizes the signal (studied distribution) to noise (novel distribution) ratio for each criterion (see, e.g., Snodgrass et al., 1985, pp. 68–70). It is important to have a measure for criterion because criterion can vary across individuals and even within the same individual.

However, the criterion measures of signal detection theory do not capture all changes in confidence ratings. In particular, they miss changes in confidence ratings at the same criterion. Criterion is held constant by using a forced-choice recognition test (see Green & Swets, 1974, chap. 2), yet confidence increased when my subjects studied a picture that was related to the target. In a yes/no recognition

Table 3
Median Confidence Ratings for Correct and Incorrect Choices
as a Function of Control/Experimental Condition

Study	N	Correct Choices		Sign Test $p <$	Incorrect Choices		Sign Test $p <$
		Control	Experimental		Control	Experimental	
A	23	4.60	4.66	.417	3.00	4.00	.105
B	22	3.82	4.50	.143	2.00	2.50	.067
C	18	3.69	3.84	.500	1.29	2.63	.004*
D	18	4.75	4.50	.119	1.50	2.50	.004*
E	34	4.47	4.53	.196	2.00	3.25	.020*
F	24	2.50	2.69	.032*	2.00	2.73	.001*
G	24	3.80	4.14	.032*	2.54	3.54	.154
H	18	3.79	4.25	.079	1.83	2.67	.002*
I	18	4.03	4.44	.009*	2.00	2.94	.004*
J	17	4.67	5.20	.315	2.50	2.25	.315
K	16	5.52	5.07	.815	2.09	3.50	.038*
L	34	4.79	5.00	.304	2.18	2.50	.304
M	17	3.67	4.30	.315	1.75	2.75	.025*
N	18	3.96	4.21	.174	2.86	3.00	.015*

Note—Confidence was rated on a Likert scale, from *guessing* (1) to *certain* (6). *Confidence was higher in the experimental condition than in the control condition ($p < .05$).

test it is common to compare beta for two conditions, but it is not common to report the median confidence ratings for two conditions. The increase in confidence ratings can only be appreciated by directly comparing confidence ratings for the two conditions.

The dissociation between accuracy and confidence suggests that confidence is not determined solely by specific details about the target (or memory strength). Specific details about the target must contribute to confidence ratings, because the subjects were more confident of their accurate choices than they were of their inaccurate choices. However, the dissociation between accuracy and confidence indicates that additional factors must contribute to confidence.

A Dissociation Between Memory for Details and Memory for the General?

Why did seeing the related picture increase confidence without increasing accuracy? One possibility is that accuracy depends on memory for specific details of the target, whereas confidence is influenced by a "generic" memory for the scene. In the modified recognition test, the target must be discriminated from the novel alternative on the basis of memory for specific details. Memory for target details may be compromised when the subject accesses traces of related events (see Chandler, 1991, 1993). In contrast, confidence may be influenced by a "generic" memory for what is common across events (see MacLeod & Nelson, 1976, pp. 132-133; see also Watkins & Kerker, 1985, for the distinction). Seeing the related picture would make the scene more familiar and would boost confidence.

The present study is not the first to suggest that metamemory can be influenced by the familiarity of the retrieval cues, rather than determined solely by the specific information about the target. Others have made the same suggestion for *prospective* measures of memory, including FOK judgments and TOT ratings (e.g., Costermans, Lories, & Ansay, 1992; Koriat & Lieblich, 1977; Metcalfe et al., 1993; Reder & Ritter, 1992). The results from the present study suggest that cue familiarity also influences confidence ratings, which are *retrospective* judgments regarding accuracy.

The result is somewhat counterintuitive. Effects of cue familiarity would be expected for prospective measures (e.g., FOK judgments and TOT ratings) because there is so little information about the target. FOK judgments are made rapidly, before the target can be retrieved (see, e.g., Reder, 1988, pp. 246-247), or are made after the subject has failed to recall the target (as are TOT states). In contrast, confidence ratings are made *after* a response has been made. In theory, confidence ratings could be based solely on memory for the target's details.

The results from the present studies suggest that target information influenced confidence ratings, which tended to be higher for correct choices than for incorrect choices. Nevertheless, a variable that often decreased accuracy (seeing a related picture) increased confidence. Comparing the present results with previous work suggests that the influence of cue familiarity on measures of metamemory (FOK

judgments, TOT states, and confidence ratings) is quite general.

Why would greater familiarity with the scene increase confidence in a choice that requires memory for specific details? The subjects were clearly instructed to rate their confidence that they had chosen the correct picture. If the subjects were following the directions, then they must have mistaken this general familiarity with their memory for specific details. Peterson and Pitz (1988) have noted that people believe that having more information increases accuracy, even when it does not. Perhaps my subjects realized that the scene was familiar and inferred that their memory for the details must be accurate as well. Or perhaps it is difficult to separate a general sense of familiarity from memory for the specific. Considering only the specific details may require special instructions and effort (see, e.g., D. S. Lindsay, in press), if it is possible at all.

Another Dissociation Between Accuracy and Confidence

A dissociation between confidence and accuracy shows a memory illusion. A variable influences accuracy, yet the subject is not aware that it does so. Other variables may produce similar dissociations. For example, requiring people to judge when an event occurred could produce a dissociation between confidence and accuracy. Accuracy would be lower in tasks that require people to judge when or where an event occurred. If people do not realize when they are making a setting misattribution, they will not adjust their confidence. Jacoby et al. (1989) have suggested that people are confident if they remember any specific details, even if the details refer to the nontarget event. In other words, people will misattribute the nontarget to the wrong setting (see, e.g., D. S. Lindsay, in press), without realizing their error.

Jacoby et al.'s (1989) claim is supported by some additional data collected from Studies A, B, F, and G. The subjects who saw the target (Lake A) and a related picture (Lake B) were asked to discriminate the target from either a *novel* alternative (Lake A vs. *Lake C*) or from the *nontarget* (Lake A vs. *Lake B*). Accuracy was lower when the nontarget was the alternative because the test forced the subjects to decide which picture they saw first (see Chandler, 1989). If subjects are aware that their accuracy depends on the requirements of the test, then they should be less confident in choosing the *nontarget* than in choosing the *novel* item. But the reverse was true. Median confidence ratings were higher when the incorrect choice was the nontarget item rather than the novel item (sign test, $z = 4.23, p < .01$). Confidence ratings were higher when the subjects remembered specific details (the nontarget), even when they misattributed the nontarget to the wrong time.¹

The finding argues against a relative strength model, which has recently been proposed by Marsh and Bower (1993, pp. 685-687) to account for item occurrence and source attribution judgments. Their model assumes that items vary in strength and that subjects set two criteria—

they use a lenient criterion to judge whether an item occurred, and a strict criterion to make source attributions. According to the model, both accuracy and confidence should be lower when the forced-choice test requires source attributions. Contrary to the model, a test that required source attributions decreased accuracy, but increased confidence.

These dissociations between accuracy and confidence have practical (as well as theoretical) relevance. People use confidence to gauge the accuracy of their own recollections. The present study shows that individuals tend to be more confident of their correct choices than they are of their incorrect choices. However, showing a related event often decreases accuracy while it boosts confidence. Although increasing confidence in correct choices is beneficial, increasing confidence in incorrect choices gives us the illusion that our memory is accurate.

REFERENCES

- CHANDLER, C. C. (1989). Specific retroactive interference in modified recognition tests: Evidence for an unknown cause of interference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 256-265.
- CHANDLER, C. C. (1991). How memory for an event is influenced by related events: Interference in modified recognition tests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 115-125.
- CHANDLER, C. C. (1993). Accessing related events increases retroactive interference in a matching recognition test. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 967-974.
- COSTERMANS, J., LORIES, G., & ANSAY, C. (1992). Confidence level and feeling of knowing in question answering: The weight of inferential processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 142-150.
- GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics*. New York: Wiley.
- JACOBY, L. L., KELLEY, C. M., & DYWAN, J. (1989). Memory attribution. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 391-422). Hillsdale, NJ: Erlbaum.
- KORIAT, A., & LIEBLICH, I. (1977). A study of memory pointers. *Acta Psychologica*, **41**, 151-164.
- LEHMANN, E. L. (1975). *Nonparametrics: Statistical methods based on ranks*. San Francisco: Holden-Day.
- LINDSAY, D. S. (in press). Memory source monitoring and eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments*. New York: Cambridge University Press.
- LINDSAY, R. C. L., WELLS, G. L., & RUMPEL, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations? *Journal of Applied Psychology*, **66**, 79-89.
- 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.
- LUUS, C. A. E., & WELLS, G. L. (1992, June). *Social influences on eyewitness confidence*. Poster presented at the meeting of the American Psychological Society, San Diego, CA.
- MACLEOD, C. M., & NELSON, T. O. (1976). A nonmonotonic lag function for false alarms to associates. *American Journal of Psychology*, **89**, 127-135.
- MARSH, R. L., & BOWER, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 673-688.
- METCALFE, J., SCHWARTZ, B. L., & JOAQUIM, S. G. (1993). The cue familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 851-861.
- NELSON, T. O., DUNLOSKY, J., WHITE, D. M., STEINBERG, J., TOWNES, B. D., & ANDERSON, D. (1990). Cognition and metacognition at extreme altitudes on Mount Everest. *Journal of Experimental Psychology: General*, **119**, 367-374.
- NELSON, T. O., MCSPADDEN, M., FROMME, K., & MARLATT, G. A. (1986). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General*, **115**, 247-254.
- NELSON, T. O., & NARENS, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-173). New York: Academic Press.
- PETERSON, D. K., & PITZ, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 85-92.
- REDER, L. M. (1988). Strategic control of retrieval strategies. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 227-259). New York: Academic Press.
- REDER, L. M., & RITTER, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 435-451.
- RYAN, R. H., & GEISELMAN, R. E. (1991). Effects of biased information on the relationship between eyewitness confidence and accuracy. *Bulletin of the Psychonomic Society*, **29**, 7-9.
- SHIMAMURA, A. P., & SQUIRE, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 452-460.
- SHIMAMURA, A. P., & SQUIRE, L. R. (1988). Long-term memory in amnesia: Cued recall, recognition memory, and confidence ratings. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 763-770.
- SNODGRASS, J. G., LEVY-BERGER, G., & HAYDON, M. (1985). *Human experimental psychology*. New York: Oxford University Press.
- TULVING, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press, Clarendon Press.
- WATKINS, M. J., & KERKAR, S. P. (1985). Recall of a twice-presented item without recall of either presentation: Generic memory for events. *Journal of Memory & Language*, **24**, 666-678.
- WELLS, G. L., FERGUSON, T. J., & LINDSAY, R. C. L. (1981). The tractability of eyewitness confidence and its implications for triers of fact. *Journal of Applied Psychology*, **66**, 688-696.

NOTE

1. A similar effect was reported by Loftus, Miller, and Burns (1978, p. 27) and by Ryan and Geiselman (1991). However, both studies suggested that a nontarget (a yield sign) was shown instead of the target (a stop sign). The procedure introduces social demands to choose the suggested item with great confidence. See Luus and Wells (1992) for the effects of social demands on confidence ratings.

(Manuscript received April 2, 1993;
revision accepted for publication September 28, 1993.)