

Some deleterious consequences of the act of recollection

JONATHAN W. SCHOOLER, RACHEL ANN FOSTER, and ELIZABETH F. LOFTUS
University of Washington, Seattle, Washington

Two experiments investigated the impact of responding to recognition test items that do not include a correct alternative. In Experiment 1, subjects who were given exclusively incorrect response alternatives were less likely than control subjects to favor the correct alternatives on a second recognition test. Analysis of subjects' responses indicated that commitments, rather than distractor familiarity, was the main source of this effect. In Experiment 2, an impairing effect of committing to an incorrect alternative was observed even when the initial distractors were excluded from the final test. Thus, this decreased performance cannot simply be attributed to a bias toward remaining consistent. One interpretation of these results is that committing to a distractor causes subjects to remember a false detail that can interfere with their later ability to access the original information. Other potential theoretical and applied implications of these results are explored.

Over the years considerable research has been devoted to determining the factors that impair memory recollection. Expectations, interference, postevent suggestions, and time itself have all been shown to change the way we recall events (see Loftus, 1979). Another activity that has potentially detrimental effects on memory reports is the act of recollection. Although ample research has assessed the improvement in final memory accuracy that results from multiple earlier recollections, much less attention has been focused on determining the conditions under which the act of recollection may impair subsequent performance (for reviews, see McDaniel & Masson, 1985; Richardson, 1985).

An impairing effect of recollection may be expected when subjects initially recollect inaccurately. There are two types of errors in an initial memory test that may impair performance in a subsequent memory test: omissions (failing to recall an item) and intrusions (recalling a non-memorized item). The impairing influence of memory omissions has been used to account for why a recall test can reduce performance on subsequent recognition tests. For example, several investigators (e.g., Brown & Packham, 1967; Postman, Jenkins, & Postman, 1948) have suggested that one effect of an interpolated recall test before a recognition test is to cause "unlearning" of those items that were not initially recalled. Although subsequent research has questioned whether omissions actually

produce unlearning (Broadbent, 1973), it is well documented that the act of recollection can cause subjects to recognize previously recalled items at the expense of omitted ones (for further details, see Cooper & Monk, 1976; Richardson, 1985).

The impact of inaccurate memory intrusions on subsequent recollection is less well understood. Research suggesting that inaccurate intrusions may impair subsequent recollection dates at least back to Bartlett (1932). Bartlett had subjects repeatedly reproduce simple line drawings that they had studied. He observed a "perseveration of errors." Errors that appeared after the first recollection tended to be maintained in subsequent recollections. Kay (1955), using a slightly different paradigm, observed an even more striking example of error perseveration. Kay repeatedly asked subjects to recall a previously heard passage. After the subjects recalled the passages as best they could, Kay presented them again in order to help the subjects correct their recollections. To his surprise, Kay observed that despite the multiple presentations of the correct text, the subjects continued to produce their initial errors. It appeared that the act of recalling inaccurate details made subjects' memories resistant to subsequent corrections. Kay's results have been replicated using more precise methods of analysis, and have been generalized to a recognition paradigm (Howe, 1970, 1972).

Although the results of Bartlett (1932), Kay (1955), and Howe (1970, 1972) are suggestive in demonstrating the "freezing effects" of recalling inaccurate information, they are still inconclusive because of the lack of appropriate control groups. Because of this methodological flaw, these studies did not demonstrate that subjects would have performed better had they not recalled the inaccurate facts. In short, demonstrating the negative consequences of inaccurate recollections requires comparing the performance of subjects who actively recollect some information to that of subjects who do not engage in recollection.

An earlier version of this paper was presented at the Annual Meeting of the British Psychological Society, May 1986. This research was supported by a grant to Elizabeth Loftus from the National Institute of Mental Health and by a Graduate Opportunity Program Research Assistantship and a National Science Foundation Minority Graduate Fellowship awarded to Rachel Ann Foster. William Benzing helped to develop the betting-form technique and method of analysis. We are grateful to Debra Peters, Karsten Moen, and William Bissell for running subjects in Experiment 2. Please address all correspondence to Jonathan W. Schooler, who is now at the Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA 15260.

One difficulty in investigating the influence of inaccurate recollections stems from the previously noted observation that asking subjects to repeatedly recall learned information often tends to improve performance. The additional rehearsals appear to increase the number of traces potentially available for accessing the information (McDaniel & Masson, 1985). Thus, when multiple recollections are compared with a single recollection, the impairing effects of incorrect recollections may be masked by the otherwise facilitating effects of correct recollections. One way to isolate the impairing effects of inaccurate recollections is to provide subjects with test alternatives that do not include the correct response; this virtually guarantees that the memory report will be incorrect. If the inaccurate recollection impairs subsequent performance, then subjects who are only given incorrect response alternatives on the first test should subsequently perform worse than subjects who are not drawn into an inaccurate recollection.

A few studies have explicitly investigated the effects on performance of providing subjects a recognition test with no correct alternatives. In a study by Brown, Deffenbacher, and Sturgill (1977), subjects interacted with two individuals and later were asked to identify them in a photo lineup that did not include a photo of one of the individuals. A few days later the subjects were given a second recognition test that included photos of both individuals, as well as a distractor photo from the earlier test. In this final test, the subjects were as likely to recognize the previously seen distractor photo as they were to recognize the target photo they had not seen in the earlier test. These results suggest that participating in a recognition test that does not include a correct response may cause subjects to later favor an incorrect alternative. However, this effect does not necessarily reflect the influence of inaccurate recollection. It is also possible that the multiple presentations of the distractor photo made the distractor more familiar and thus more likely to be misidentified.

Underwood and Freund (1970) documented the increase in false alarms that can result from repeatedly presenting distractors. In several experiments, subjects learned word lists and were later given a two-alternative recognition test. The test was devised so that the same distractor could appear in one, two, or three different test pairs. Subjects were more likely to incorrectly recognize distractors that had appeared more than once. Repeated presentation apparently increased the distractors' familiarity and thus their likelihood of being misidentified. Given this observation, it seems plausible that the subjects in Brown et al.'s (1977) study also might have favored the previously seen distractor photo because of the increased familiarity that resulted from multiple presentations.

In short, there are two possible ways in which taking a test without correct alternatives may affect subsequent performance: (1) repeatedly seeing the distractors may make them more familiar, and thus more likely to be selected on a final test, and (2) the act of committing to an incorrect response may predispose an individual toward

that particular distractor. If increased familiarity is the major source of this effect, subjects should prefer previous distractors to new distractors, regardless of whether or not they had previously selected them. Alternatively, if commitment is the critical factor, subjects should favor only the distractor that they had previously selected.

Gorenstein and Ellsworth (1981) conducted a study similar to that of Brown et al. in which they explicitly examined the effect of commitment. Gorenstein and Ellsworth staged an incident and later gave half of the witnesses a face recognition test that did not include a correct alternative. Subjects who were induced to recognize incorrect distractor photos performed worse on the subsequent test than did control subjects who were not given the interpolated test. Furthermore, the nature of the errors supported a commitment effect: subjects were most likely to identify the distractor that they had previously selected. The distractors that had been previously seen but not selected were no more likely than new distractors to be chosen on the final test. Thus, Gorenstein and Ellsworth's findings supported a commitment effect while providing no evidence for an effect of familiarity.

Why would committing to an incorrect response cause subjects to subsequently prefer their earlier choice over the correct alternative? One possibility is that the act of incorrect commitment causes subjects to inaccurately recollect the incident. This distorted act of recollection then interferes with subjects' subsequent memory performance, so that they later remember their earlier selection rather than the original face. Alternatively, a commitment effect may have nothing to do with the act of recollection, but may instead simply reflect an attempt to remain consistent. Before further discussion of the possible effects of the act of commitment, however, it should be pointed out that presently the evidence for a commitment effect is rather weak.

There are a number of reasons for caution in drawing firm conclusions from Gorenstein and Ellsworth's (1981) results. First, they had only one critical item and very few subjects (18 in each of two conditions). There were only small differences between the numbers of correct subjects in their various conditional analyses. For example, the difference between the numbers correct in the control and interpolated conditions was only 5 subjects. The conclusion that commitment was the central source of this effect was also based on a difference of 5 subjects. Given these small numbers, it seems premature to assume that committing to an incorrect alternative will subsequently bias subjects against the correct alternative.

A second limitation of Gorenstein and Ellsworth's (1981) study was the limited power of their study to evaluate the effects of distractor familiarity. Because most errors in the multiple-test condition involved committing to the previously selected distractor, the few remaining errors made it rather difficult to assess whether there was any residual effect of familiarity, that is, whether subjects preferred previously seen but not selected distractors over new distractors. A more sensitive test of the

effects of familiarity would be to give subjects an opportunity to indicate the probability that each of their non-preferred alternatives could be correct. This "betting-form" procedure would permit every subject to indicate the degree to which distractor familiarity affected his/her response.

Interpretation of these earlier studies examining the effects of inaccurate recognition tests is further clouded by another concern. Specifically, in the previous studies the initial incorrect alternatives were present in the final test, thus presenting two alternative explanations for the effect of the initial incorrect alternatives: (1) the subjects' memory for the original information might have been impaired, or (2) the subjects might simply have been biased toward selecting one of the previously presented incorrect responses. A more suitable test for determining whether inaccurate recollection truly impairs memory involves examining whether accuracy is reduced even when the initial incorrect distractors are omitted from the final test. If performance on a final recognition test is reduced even when the initial distractors are excluded, then the reduced performance cannot be due simply to a bias toward the initial distractors. Consequently, memory impairment is implicated.

The current series of experiments was designed to further explore the potentially impairing effects of inaccurate recollection on subsequent eyewitness performance. Table 1 presents the basic design of Experiments 1 and 2. In Experiment 1 we attempted a more rigorous examination of the issues explored by Gorenstein and Ellsworth (1981), using more subjects, a greater variety of test items, and a betting-form memory test to more sensitively assess the possible influence of distractor familiarity. In Experiment 1 the subjects viewed a series of slides containing some critical details (e.g., the subjects saw item A). Later, subjects in the two recollect conditions were given a recognition test that included several critical questions. In the distractor-distractor recollect condition, the critical questions included two distractors but did not include the correct alternative (e.g., "Did you see item B or item C?"). In the target-distractor recollect condition, the critical

questions included the correct alternative and a distractor (e.g., "Did you see item A or item B?"). In the no-recollect condition, the subjects were given an unrelated filler task instead of a recognition test. In the final test, the subjects were given test alternatives that included the original target items, the distractors appearing in the earlier recognition test, and a new distractor (e.g., "Did you see item A, item B, item C, or item D?"). Rather than forcing the subjects to choose a single option, this final test allowed them to divide probability points among the options. The betting-form technique thereby provided a measure of the potential influence of familiarity on test items that would not have been selected in a traditional forced test.

Experiment 1 demonstrated that commitment to a distractor on the initial recognition test was a source of subsequent errors. It was unclear, however, whether this impairment in performance was the result of a simple bias toward those earlier responses or the result of incorporation of this incorrect recollection into subjects' memories of the event. If a bias to appear consistent were the source of the commitment effect, then removing the earlier distractors from the final test should eliminate this bias and thus the influence of the earlier test. Alternatively, if the initial test caused subjects to incorrectly remember the earlier event, they might still have been affected by these false memories, even if the earlier alternatives did not appear on the final test. To assess this possibility, in Experiment 2 we used a simplified version of the same basic design as Experiment 1, with one critical exception: the previously presented distractors were excluded from the final test. In Experiment 2 the subjects viewed a slide sequence including critical items (e.g., item A) and then received a recognition test. In the recollect condition, subjects received a test that included a critical question with two distractors but no correct alternative (e.g., item B and item C). In the no-recollect condition, subjects received a questionnaire that did not mention the critical item. The final test included the original alternative and a never-mentioned distractor (e.g., item A and item D). Because the initial distractors were not included in the final test,

Table 1
Basic Design of Experiments 1 and 2

Condition	Presented	Test 1	Test 2
Experiment 1			
Distractor-Distractor Recollect	Item A	"Did you see Item B or C?"	"Did you see Item A, B, C, or D?"
Target-Distractor Recollect	Item A	"Did you see Item A or B?"	"Did you see Item A, B, C, or D?"
No-recollect	Item A	Filler activity	"Did you see Item A, B, C, or D?"
Experiment 2			
Distractor-Distractor Recollect	Item A	"Did you see Item B or C?"	"Did you see Item A or D?"
No-recollect	Item A	Filler questions	"Did you see Item A or D?"

the observation of decreased performance resulting from inaccurate recollection could not be attributed to a bias toward the initial alternatives, and would consequently favor the hypothesis of memory impairment.

EXPERIMENT 1

Experiment 1 was designed to further explore the mechanisms by which providing a recognition test with exclusively incorrect test alternatives affects subsequent performance. In Experiment 1, subjects viewed a slide sequence of a simulated wallet snatching. Recollect subjects were given a recognition test; no-recollect subjects engaged in an unrelated filler activity. Of those subjects who received the initial recognition test, some were given test alternatives that included the correct response (target-distractor recollect). The remaining subjects were given test alternatives that did not include the correct response (distractor-distractor recollect). Finally, all subjects were given a second recognition test that asked the subjects to divide 100 probability points among four alternatives: the target, the two incorrect distractors from the distractor-distractor recollect condition, and a new distractor.

Method

Subjects. The subjects were 85 undergraduate students at the University of Washington who participated for course credit. The subjects were run in groups of 10 to 15.

Design. The subjects were randomly assigned to three conditions: no-recollect ($n=32$), target-distractor recollect ($n=24$), and distractor-distractor recollect ($n=29$). In the no-recollect condition, the subjects did not take an initial recognition test. In the two recollect conditions, the subjects took an initial two-alternative forced-choice test containing four critical questions. In the target-distractor recollect condition, each critical question included the original item and a never-seen distractor as response alternatives. In the distractor-distractor recollect condition, each critical question included two never-seen distractors.

Procedure. The subjects were told that they were participating in a study of perceptual skills, and were then shown a slide sequence depicting a wallet snatching at a rate of approximately 8 sec per slide. After participating in a 25-min unrelated filler activity, the subjects received a postevent narrative that was described as a "police report," which they were to proof for spelling errors. The narrative did not mention any information on which subjects were later tested. After the postevent narrative, the subjects engaged in a 5-min unrelated filler task. This task was followed by a vocabulary test for no-recollect subjects and by one of two recognition tests (target-distractor, distractor-distractor) for recollect subjects. Immediately after completing the vocabulary test or the recognition test, the subjects were given a final recognition test, which asked the subjects to indicate for each item what they believed they actually saw. They responded by distributing 100 probability points among four alternatives: the target, the two distractors from the initial recognition tests, and a third distractor that had never been seen or mentioned before.

The betting-form measure allowed subjects to give a variety of different responses. For example, assume subjects saw item A, had been previously tested with items B and C, and selected item B. Item D they had never seen or read before. Possible point distributions for this example are as follows: The subjects could indicate total confidence in the correct alternative by assigning 100 points

to item A and 0 points to each of the remaining alternatives. They could indicate total confidence in their earlier selection by assigning 100 points to item B and 0 to each of the other alternatives. They could indicate equal confidence in the two previously seen distractors by assigning 50 points to item B and 50 points to item C. They could indicate that they had no idea which alternative was correct by assigning 25 points to each of the four alternatives. Any distribution that allocated the full 100 points among the four alternatives was allowable.

Results

Performance on the second recognition test was analyzed using two different measures. The first measure was *mean correct*, or the mean number of probability points assigned to each correct item. The second measure was *plurality correct*, or the number of critical items per subject (maximum = 4) for which the correct alternative was assigned more probability points than any other alternative. The mean correct provides a combined measure of confidence and accuracy by assessing overall confidence in the correct alternative relative to other alternatives. The plurality correct exclusively measures accuracy of performance by assessing the proportion of subjects who preferred the correct alternatives without consideration of the magnitude of their preference.

The pattern of results was the same for both types of analyses and is presented in Table 2. There was a main effect of recollection condition for both mean correct and plurality correct measures [$F(2,337) = 19.10$, $MSe = 1,535.76$, $p < .01$, and $F(2,82) = 13.46$, $MSe = .97$, $p < .01$, respectively]. A comparison between the two recollect conditions and the no-recollect condition revealed that both mean correct and plurality correct measures were greater in the target-distractor recollect condition than in the no-recollect condition [$t(222) = 2.60$, $p < .05$, and $t(54) = 2.26$, $p < .05$, respectively]. However, subjects had lower mean correct and plurality correct in the distractor-distractor recollect condition than in the no-recollect condition [$t(242) = 3.66$, $p < .01$, and $t(59) = 3.09$, $p < .01$, respectively]. In short, a comparison of the three conditions indicated that responding to a test item that included a correct alternative improved subsequent performance, whereas responding to a test item that did not include a correct alternative impaired performance.

The subjects were instructed to distribute probability points among the various alternatives; they were not told to assign a plurality of points to any alternative. The subjects could indicate that they had no preference for any alternative by assigning an equal number of probability points to each. The percentages of items for which this

Table 2
Mean Probability Points and Percentage of Items
Assigned a Plurality in Experiment 1

Condition	Mean Points (out of 100)	% Assigned a Plurality
Target-Distractor	71.10	66
Distractor-Distractor	38.86	31
No-recollect	57.81	51

occurred (i.e., the subjects were guessing) in each condition are as follows: target-distractor recollect, 24%; distractor-distractor recollect, 20%; no-recollect, 25%. It might have been expected that receiving test items with no correct alternative would affect subjects' willingness to express a preference. However, each condition had approximately the same proportion of subjects who assigned an equal number of probability points to each alternative ($F < 1$, $MSe = .97$).

A comparison of the mean number of probability points subjects assigned to each distractor in the distractor-distractor recollect condition was performed to elucidate the relatively low accuracy of subjects in this condition. The previously selected distractor received a mean of 46.59 points; the nonselected distractor, 8.13 points; the new distractor, 6.41 points. In support of a commitment effect, subjects overwhelmingly favored their previously selected distractor over the other distractors [$F(2,115) = 101.23$, $MSe = 591.59$, $p < .01$]. The results provide little evidence for an effect of distractor familiarity; the mean number of probability points assigned to the previously seen but not selected distractor was only slightly greater than that assigned to the new distractor ($t < 1$).

Discussion

Experiment 1 reliably demonstrated that participating in a recognition test that includes items with exclusively incorrect alternatives can impair subsequent performance. Furthermore, it provided strong evidence that commitment to an incorrect alternative is the major source of this effect. Subjects in the distractor-distractor recollect condition overwhelmingly preferred the distractor that they had previously selected over the other distractors. Even with the more sensitive betting-form measure, previous exposure to the nonselected distractor appeared to have little effect on subjects' responses. Subjects assigned only a slightly greater number of probability points to the nonselected previously seen distractor than to a new distractor. These results indicate that the reduced performance of distractor-distractor recollect subjects was not simply due to their previous exposure to exclusively incorrect alternatives, but was instead the result of actively committing to a particular response.

At first glance, our results appear to be in conflict with those of Underwood and Freund (1970), who reported an effect of distractor familiarity. The differences between their paradigm and ours are too great to permit any firm conclusions about this disparity. Nevertheless, our results suggest a possible reinterpretation of Underwood and Freund's observation that repeatedly presented distractors were more likely to be misidentified. They interpreted this finding as supporting a familiarity effect; that is, they concluded that multiple exposures to the distractor was the critical factor. The following reasonable assumptions suggest a rather different interpretation of this result: (1) each time a distractor is presented there is some probability that it will be misidentified, and (2) once a distractor has been misidentified, it will be more likely than

a new distractor to be subsequently misidentified (this assumption seems particularly plausible, given our current findings). Accordingly, the probability of selecting a previously seen distractor may equal the probability of misidentifying a new distractor plus some additional probability resulting from the possibility that the distractor was selected on an earlier presentation. In short, Underwood and Freund's results could also be explained in terms of a commitment effect; however, such a conclusion would require a reanalysis of their data in which misidentification of repeated distractors was conditionalized on subjects' earlier responses. The present findings suggest that such a reanalysis may be revealing.

As mentioned earlier, there are two mechanisms by which commitment may cause an impairment in subsequent performance: (1) the demand characteristics of the situation may cause subjects to stick with their earlier selection to avoid appearing inconsistent, or (2) the act of committing to an inaccurate alternative may cause subjects to falsely remember that alternative. If demand characteristics were responsible for the effect, removing the earlier distractors from the final test should eliminate the pressure to remain consistent, and thus eliminate the performance impairment. If, however, the act of committing to an incorrect response induced subjects to falsely remember that detail, then this inaccurate recollection may have interfered with subjects' ability to access the original information. If so, these subjects might have shown impaired performance even if the earlier distractors were omitted from the final test. Experiment 2 addressed this issue.

EXPERIMENT 2

It has been observed in a variety of research paradigms that the effects of an interpolated activity can depend on whether the information associated with that activity is present in the final test. For example, Broadbent (1973) observed that the impairing effects of a recall test on a subsequent recognition test were eliminated when previously recalled items were excluded from the latter test. Similarly, McCloskey and Zaragoza (1985) observed that the impairing effects on a subsequent recognition test of reading postevent suggestions were eliminated when the suggested information was excluded as a test alternative. These observations led the above authors to conclude that the interpolated activity did not genuinely impair memory for the original information. Rather, in both cases, it appeared that reduced performance on the final test was due to a bias toward information associated with the interpolated activity. In Experiment 2, we examined whether the impairing effects of receiving test questions with no correct alternative would still occur when the initial alternatives were omitted from the final test. If inaccurate recollection truly impairs memory for the original information, then items that were initially associated with incorrect test responses should be less accurately recognized than initially omitted items, even when the final test does not include the previously presented distractors.

Method

Subjects. The subjects were 121 undergraduates from the University of Washington who participated in the study for course credit.

Stimuli. The stimuli were two sequences of 38 slides depicting a burglary, in which a man enters a house, searches through various rooms, and then escapes. The two sequences differed on two critical items: one group of subjects saw a red robe and a bottle of Ivory dishwashing liquid, and the other group saw a blue robe and a bottle of Sunlight dishwashing liquid.

The initial two-alternative forced-choice recognition test included nine filler questions and one critical question. Two versions of the initial test were used: one asked about the type of soap, the other about the color of the robe. For each subject, the initial test included a question about one of the critical items (recollect items) and omitted a question about the other critical item (no-recollect items). The critical question in each test did not include a correct response alternative. For example, subjects who had actually seen Ivory soap received a test question asking them whether they had seen Dawn or Palmolive soap. The same critical questions were used for both versions of the slide sequence.

The final two-alternative forced-choice test included eight filler questions and two critical questions. The subjects were instructed to indicate for each item what they actually saw and to indicate on a 3-point scale their confidence in their decisions (1 = *guessing*, 3 = *certain*). Half (four) of the filler questions used in the earlier test were included in the final test. Each critical question included the previously seen item and a never-mentioned distractor. As a result of counterbalancing what the subjects originally saw, the same two alternatives served approximately equally often as the correct response and the never-mentioned distractor. For example, one of the critical questions asked whether the subject had seen Sunlight or Ivory dishwashing liquid. For subjects who had previously seen Ivory in the slide sequence, the "Ivory" response alternative was correct and the "Sunlight" response alternative was a never-mentioned distractor. Conversely, for subjects who had previously seen Sunlight in the slide sequence, the "Sunlight" response alternative was correct and the "Ivory" response alternative was a never-mentioned distractor. Four conditions resulted from counterbalancing the two versions of the presented slides with the two versions of the initial test: presented—blue robe and Sunlight soap, recollect—type of robe ($n=28$); presented—blue robe and Sunlight soap, recollect—type of soap ($n=30$); presented—red robe and Ivory soap, recollect—type of robe ($n=31$); presented—red robe and Ivory soap, recollect—type of soap ($n=32$).

Counterbalancing the original critical items in the manner described above eliminated potential response biases in the final test. Specifically, subjects might have taken the strategy of selecting the final test alternative that was most similar to the incorrect alternative that they had chosen in the earlier test. If one of the final distractors was consistently more similar to the earlier items than was the correct alternative, the above strategy might have biased subjects toward an incorrect response. However, since each critical alternative in the final test was approximately as often correct as incorrect, any similarities between the initial alternatives and the final alternatives should not have systematically affected subjects' responses.

Procedure. The subjects were informed that they were going to watch a slide sequence of a burglary and that they should pay particular attention to details. They viewed the slide sequence at a rate of 8 sec per slide. After a 10-min filler, the subjects received the first recognition test. Upon completion of the recognition test, the subjects were given a 5-min filler task and then the second recognition test. The filler task involved completing a crossword puzzle.

Results

On the final test, the subjects' accuracy rate was 66% for the critical items that had previously appeared on the

initial test with no correct response alternative, compared with an accuracy rate of 75% for items that had not been included on the previous test. Since each subject responded to an item from each of the two conditions, the proportions were compared using a Z test for correlated data. This analysis revealed that subjects were significantly less accurate in responding to recollect items than to no-recollect items ($Z = 1.68, p < .05$, one-tailed).

Subjects' confidence scores for their responses to critical items were also examined. A within-subjects analysis comparing recollect and no-recollect conditions conditionalized for accuracy would have been useful. Such an analysis would require that each subject provide a confidence score for each of the four response-condition contingencies: correct-recollect, incorrect-recollect, correct-no-recollect, and incorrect-no-recollect. However, such an analysis was not possible, because each subject provided a score for only two of the four response-condition contingencies. Consequently, two separate between-subjects analyses were conducted for each of the two critical items (type of soap and color of robe). This approach allowed each subject to contribute one independent confidence score to each analysis. Because the pattern of results was the same for both items, the confidence data were collapsed across items. The mean confidence scores for recollect and no-recollect items are presented in Table 3.

Although the subjects' confidence scores were slightly lower for previously presented critical items than for items that had not been previously presented, this difference was not significant [$F(1,113) < 1, MSe = .783$ (soap); $F(1,113) = 1.12, MSe = .625$ (robe)]. The subjects were significantly more confident for correct responses than for incorrect responses [$F(1,113) = 10.57$ (soap); $F(1,113) = 10.08, p < .01$ (robe)]. There was no interaction between condition and accuracy [$F(1,113) < 1$ (soap and robe)]. (Note: Some subjects did not provide confidence values, and the degrees of freedom were corrected to reflect this.)

Discussion

In Experiment 2, subjects performed more poorly on the final recognition test items when they had earlier responded to versions of those items with exclusively incorrect test alternatives. This impairment occurred even though the final recognition test items did not include the previously presented incorrect alternatives. These results indicate that the presence of exclusively incorrect alter-

Table 3
Subjects' Mean Confidence Scores for Their Responses
to Critical Items in Experiment 2

Type of Response	Condition		Overall
	Recollect	No-recollect	
Correct	2.08	2.28	2.19
Incorrect	1.46	1.52	1.49
Overall	1.89	2.10	

Note—1 = *guessing*, 3 = *certain*.

natives on a recognition test does not simply cause subjects to prefer those alternatives on a subsequent test; rather, the act of committing to an incorrect alternative appears to interfere with subjects' subsequent ability to recognize the correct alternative.

It was possible that changes in confidence caused the decreased accuracy exhibited by subjects who had been previously presented with exclusively incorrect test alternatives. Recent studies have suggested that negative feedback can reduce confidence in memories and the consequent willingness to rely on one's memories (e.g., Gudjonsson & Clark, 1986). Accordingly, the lack of a correct alternative on the initial test could have served as a source of negative feedback, causing some subjects to feel generally confused and to lose confidence in their memories. During the second test, the reduced confidence of these subjects could have caused them to guess indiscriminately. This notion leads, however, to the prediction that subjects should be significantly less confident after being exposed to questions without correct answers. Such lowered confidence was not observed. Thus, a change in memory confidence does not appear to account for the present results. Instead, a more compelling explanation seems to be that subjects' access to the originally encoded information was impaired as a result of previously committing to an incorrect alternative.

Although memory impairment is a feasible hypothesis, it remains possible that inaccurate recollections may have other consequences as well. For example, in Experiment 1, in which the initial incorrect alternatives were included in the final test, it is possible that some subjects might have experienced impaired memory for the original alternative, whereas other subjects simply might have been biased toward previously selected alternatives. In short, although Experiment 2 suggests that recognizing incorrect alternatives can impair subjects' original memories, it does not rule out other ways in which exclusively incorrect test alternatives may reduce subjects' subsequent recognition accuracy.

GENERAL DISCUSSION

Collectively, the results of the two experiments indicate that participating in a recognition test that encourages incorrect responses can reduce subjects' accuracy on a subsequent test. In Experiment 1, we showed that commitment was an important factor in mediating this effect. In Experiment 2, we observed that memory performance was lowered even when the final test did not include the earlier distractors. The latter result suggests that committing to an incorrect alternative does not simply bias subjects toward their earlier responses.

What is it about the act of committing to an incorrect alternative that produces this decrement in subsequent performance? One possibility is that the effects of incorrect commitment may be the product of demand characteris-

tics associated with the final test. Accordingly, the inclusion of the previous distractors may compel subjects to appear consistent and consequently to disregard the correct alternative in favor of their earlier selections. This hypothesis cannot explain the entire effect of commitment, however, because our subjects were still affected by their earlier incorrect responses even when those selections were not presented. A second possibility is that subjects lose confidence in their memories and simply feel compelled to guess. In both experiments, however, we observed no difference in guessing rates. A final possibility is that the act of committing to an incorrect response causes subjects to falsely remember that information. This inaccurate recollection and the false memory it produces later cause interference that impairs the subjects' ability to remember the original details. Although we can never be sure that when subjects commit to an incorrect response they are engaging in a genuine act of inaccurate recollection or that a false memory is produced, the present results are consistent with this interpretation.

The present findings may appear, at first glance, to conflict with those of two studies reported by Mandler and his colleagues. Mandler, Pearlstone, and Koopmans (1969) presented subjects with a word list followed by an immediate recall test. The subjects were then given a recognition test that included synonyms for the original words and unrelated distractors. Finally, the subjects were given a second recall test. Despite the interpolated recognition test with exclusively incorrect alternatives, the subjects actually performed better on the subsequent recall test than they had initially. This study was later replicated using homophones, rather than synonyms, in the interpolated recognition test (Mandler, Meltzer, & Pearlstone, 1969).

There are a number of reasons why Mandler and his colleagues may have observed an effect of inaccurate recollection different from the effect we observed. First, Mandler used word learning, which may not generalize to more naturalistic eyewitness situations. Second, the distractors in Mandler's studies were similar to the correct alternatives. When Mandler and his colleagues used a recognition task with unrelated fillers, no facilitation was observed. Third, the critical dependent measure on the final test in Mandler's experiments was recall, rather than recognition. It is possible that the effects of inaccurate recollection may depend on the form of the final test. For example, errors of omission are frequent in recall tests but are not possible in recognition tests. In the case of a recall test, in which subjects may omit items for which they have some memory, exposing subjects to related distractors on an interpolated recognition test may prompt subjects to recall otherwise "forgotten" items. However, in the case of a recognition test, on which errors of omission cannot occur, exposure to related distractors is unlikely to prompt subjects to recall otherwise forgotten items. The present results would lead us to expect, in-

stead, that related distractors would impair performance. It would be useful for future research to explicitly compare the influence of inaccurate recollection on subsequent recall and recognition tests.

Our observation of recognition impairment in the present study also contrasts with a recent similar investigation in which subjects were given an interpolated activity that included incorrect information. As previously mentioned, McCloskey and Zaragoza (1985) presented subjects with a misleading narrative that incorrectly referred to items seen in an earlier slide sequence. On a subsequent recognition test, the subjects were unaffected by the misleading narrative when the suggested version of an item was excluded as a test alternative. Our findings would lead us to expect that reading an inaccurate narrative would induce subjects to inaccurately recollect the slide sequence and consequently to perform less well on the final test. Two differences between our paradigm and that of McCloskey and Zaragoza may be responsible for this different pattern of results: the number of incorrect interpolated alternatives, and whether or not subjects had to explicitly commit to an incorrect alternative. In McCloskey and Zaragoza's interpolated activity, each critical item was associated with a single incorrect alternative; in the present experiments, each critical item was associated with two incorrect alternatives. In theory, this difference could be important. However, the subjects' performance in our Experiment 1 strongly indicated that they were not affected by the nonselected interpolated alternatives. It seems unlikely, therefore, that the sheer number of incorrect references to the critical item was the important factor. Rather, as suggested by the subjects' preferences for their previously selected distractors, it seems to be the act of commitment that is critical in producing memory impairment. Accordingly, by actively committing to an incorrect response, a subject may be more likely to incorporate that fact and thus experience subsequent memory interference. The relative influence of actively committing to an incorrect response, as opposed to passively reading an inaccurate postevent narrative, might be a profitable subject for future research.

The distinction between inaccurate recollection produced by reading misinformation in a postevent narrative, and responding to test questions that exclude the correct alternative, parallels in some ways the distinction between passive and active learning. Slamecka and Graf (1978) observed that subjects were more likely to subsequently recall a response they had generated than to recall the same information passively heard. These authors suggested that self-generated items were more memorable than passively received items. Admittedly, deciding between two incorrect test alternatives does not represent a completely self-generated inaccurate response; nonetheless, the fact that such a response requires a decision may make it more "self-generated" than simply reading an inaccurate detail. Therefore, the present study expands upon Slamecka and Graf's notion by suggesting that self-

generated inaccurate responses may impair memory more than does passively acquired inaccurate information. Clearly, further research is necessary to more fully explore this possibility.

What would happen if subjects were not forced to select one of the two inaccurate alternatives? What if subjects had the option of indicating that neither response was correct? Including a "none of the above" response might have eliminated the implicit misinformation associated with not including a correct alternative. Moreover, with such a procedure, subjects would not have been forced to make a commitment. However, various lines of research have indicated that subjects prefer to make positive responses. For example, in a study of face identification, up to 88% of subjects were willing to attempt an identification in a lineup in which the suspect was not actually present (Kohnken & Maass, *in press*). It seems plausible, therefore, that in the present study many subjects might have committed to an incorrect response even if a "none of the above" response had been included. Under these circumstances, the subjects would not have been compelled to select an incorrect alternative; hence, those who did so might have felt more personally invested in their responses and consequently might have experienced greater memory impairment. Thus an experimental manipulation that included a "none of the above" alternative might also produce results that would advance our understanding of the mechanisms by which inaccurate reporting impairs subsequent memory.

The present results have potentially important practical implications. Witnesses are often asked to make recognition decisions between alternatives that may not be correct. A common practice in police interrogations is to prompt witnesses when they have trouble recalling certain details. For example, if a witness cannot recall the color of a perpetrator's eyes, he/she may be asked, "Did the suspect have brown or blue eyes?" The present results suggest that such questions might possibly impair a witness's subsequent memory performance. In the above example, if the suspect actually had green eyes, then the interrogator's questions might unwittingly cause the witness to engage in an inaccurate act of recollection, and thus impair his/her ability to later recollect the correct color. Accordingly, the present research suggests that interviewers should take care not to, even inadvertently, induce subjects to decide between alternatives that may conceivably be incorrect.

Investigations of the potentially impairing influence of multiple recollections suggest that memory researchers may face a problem that has confronted physicists for some time: examining certain aspects of a phenomenon can interfere with the ability to examine other aspects of that phenomenon. For example, whether a physicist measures light as particles or waves affects how the light appears to behave. As in the case of the physics of light, the method of observing memory can affect its "appearance." As Tulving (1984) noted, "traces have no strength

independently of conditions in which they are actualized: any given trace can have many different 'strengths' depending on its retrieval conditions" (p. 233). If what we remember depends on how our memories are tested, and if how our memories are tested can affect what we can subsequently recollect, then memory researchers may never be able to know what information was potentially available to a given individual. Accordingly, tests that optimize the recovery of certain details may at the same time impair the recovery of others. We may be faced with another of those infuriating paradoxes of measurement: the act of observing a memory may change the very memory that we are attempting to observe.

REFERENCES

- BARTLETT, F. C. (1932). *Remembering*. London: Cambridge University Press.
- BROADBENT, D. E. (1973). *In defense of empirical psychology*. London: Methuen.
- BROWN, E., DEFFENBACHER, K., & STURGILL, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology*, *62*, 311-318.
- BROWN, J., & PACKHAM, D. W. (1967). The effect of prior recall on multiple response recognition. *Quarterly Journal of Experimental Psychology*, *19*, 356-361.
- COOPER, A. J. R., & MONK, A. (1976). Learning for recall and learning for recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 131-156). Chichester: Wiley.
- GORENSTEIN, G., & ELLSWORTH, P. (1981). Effect of choosing an incorrect photograph on a later identification by an eyewitness. *Journal of Applied Psychology*, *65*, 616-622.
- GUDJONSSON, G. H., & CLARK, N. K. (1986). A theoretical model of interrogative suggestibility. *Social Behavior*, *1*, 83-104.
- HOWE, M. J. A. (1970). Repeated presentation and recall of meaningful prose. *Journal of Educational Psychology*, *61*, 214-219.
- HOWE, M. J. A. (1972). Repeated presentation and retention of meaningful information. *Psychological Reports*, *31*, 840-842.
- KAY, H. (1955). Learning and retaining verbal material. *British Journal of Psychology*, *46*, 81-100.
- KOHNKEN, G., & MAASS, A. (in press). Eyewitness testimony: False alarms on biased instructions? *Journal of Applied Psychology*.
- LOFTUS, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- MANDLER, G., MELTZER, R. H., & PEARLSTONE, Z. (1969). *The structure of recognition: Effects of list tags and of acoustic and semantic confusion* (Tech. Rep. No. 7). San Diego: University of California, Center for Human Information Processing.
- MANDLER, G., PEARLSTONE, Z., & KOOPMANS, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning & Verbal Behavior*, *8*, 410-423.
- MCCLOSKEY, M., & ZARAGOZA, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*, 1-16.
- MCDANIEL, M. A., & MASSON, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 371-385.
- POSTMAN, L., JENKINS, W. O., & POSTMAN, D. L. (1948). An experimental comparison of active recall and recognition. *American Journal of Psychology*, *61*, 511-519.
- RICHARDSON, J. T. E. (1985). The effects of retention tests upon human learning and memory: An historical review and an experimental analysis. *Educational Psychology*, *5*, 85-114.
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 592-604.
- TULVING, E. (1984). *Precis of elements of episodic memory*. *Behavioral & Brain Sciences*, *7*, 223-238.
- UNDERWOOD, B. J., & FREUND, J. S. (1970). Testing effects in the recognition of words. *Journal of Verbal Learning & Verbal Behavior*, *9*, 117-125.

(Manuscript received September 16, 1986;
revision accepted for publication October 26, 1987.)