

Eyewitness memory and the importance of sequential information

NEAL E. A. KROLL, KEITH H. OGAWA, and JAMES E. NIETERS
University of California, Davis, California

Misleading postevent information increases the probability of incorrectly recognizing a detail suggested by the misleading information. This often-reproduced finding has been interpreted as demonstrating both memory impermanence (e.g., Loftus & Loftus, 1980) and recoding (e.g., Tulving, 1983). However, recent evidence suggests that postevent information affects not the memory of the original event, but rather the guessing bias when memory fails. An experiment is presented that supports this response-bias interpretation. Providing witnesses with the original sequence information, even after they had already chosen the incorrect detail on an earlier test, greatly enhanced the probability of their retrieving the original memory—previously believed to be irretrievable—and improved the validity of the witnesses' confidence ratings.

Brigham and Wolfskeil (1983) found that prosecutors and law officers regard eyewitness identification as accurate, despite repeated demonstrations to the contrary since Stern (1904). More recently, an experimental paradigm developed by Loftus (1975) has dominated the study of eyewitness memory. It is widely used to document the distorting effect of misleading questions, but researchers disagree as to the cause and locus of such distortions.

In Loftus's (1975) paradigm, subjects view slides depicting an event, then receive postevent information, typically in questions about the event, followed by a forced-choice recognition test. In the misleading information (MI) condition, a question about a critical event slide references a detail that was not present, but that will be in an incorrect alternative in the recognition test. In the no information (NI) control condition, the question does not refer to the detail differentiating the correct slide from its foil. For example, if a critical slide showed a body with a belt around the neck, a NI question asks about the scene without mentioning a belt. A MI question, however, makes an incidental reference to "a rope around the neck." The two alternatives seen in the test would then differ only with respect to this detail—the correct slide showing a belt, the foil containing a rope.

Typically, subjects given MI are more likely to choose the foil than are those given NI (e.g., Loftus, Miller, & Burns, 1978). Four explanations have been offered to explain this effect; the first three were discussed by Loftus and Loftus (1980). The *coexistence hypothesis* assumes that the original information (OI) is retained in memory, but postevent MI causes interference, making OI memory less accessible. The *demand characteristic hypothesis* assumes that the accessibility of OI is not affected; the foil is chosen to comply with perceived situational demands. Loftus and Loftus rejected these explanations, arguing instead for the *alteration hypothesis*, which assumes that MI replaces memory of OI. They concluded that information originally stored in memory was "irrevocably destroyed" (p. 409).

The coexistence hypothesis, however, does have defenders. Bekerian and Bowers (1983) found that misled subjects per-

formed worse than controls only when test slides were presented in a different sequence from that of the event slides. They argued that MI does not replace OI, but merely makes OI more difficult to access. The combination of MI interference and lack of sequence information leads to increased errors. If sequence information is supplied in the recognition test, subjects overcome the interference and retrieve OI.

These experiments provide evidence against memory replacement during presentation of MI (Cole & E. F. Loftus, 1979). However, they do not address the possibility that it occurs during the recognition test (E. F. Loftus & Schooler, 1984). According to this hypothesis, OI and MI coexist until a decision is made. If a subject—consciously or not—chooses a foil depicting MI, then this incorrect slide is remembered as what had been seen. This would explain experiments in which no evidence of OI memory was found after a misled subject had chosen the foil. However, Kroll and Timourian (1986) presented evidence that event memory can be improved *after* a misled subject chooses a foil.

A fourth explanation of the MI effect—*response bias*—is from McCloskey and Zaragoza's (1985; Zaragoza, McCloskey, & Jamis, 1987) contention that Loftus's procedure does not necessitate a postevent MI effect on OI memory. The effect may be from biasing the guessing of those who would not have remembered OI even without MI. Clearly, any demonstration of poorer memory performance—or a demonstration of later improved memory—needs to separate effects on memory from those on guessing bias.

In a previous experiment (Kroll & Timourian, 1986), subjects saw slides depicting a walk through town and an observation of a purse snatching. A questionnaire included three questions about critical slides: two gave MI and one gave correct information (CI) favoring the target over the foil. There were two recognition tests. Prior to the second test, subjects were told that the questions may have led to errors in the first test and they should try to improve their scores. Between tests, subjects "returned to the crime scene" by viewing slides of the town in the same sequence, but without the characters in the original scenario. There were also two control conditions. In one, the "return" slides were presented between tests, but in random order. In the other, the return slides were in the original order, but were shown prior to the first recognition test.

An earlier version of this report was presented in 1987 to the Second International Conference on Practical Aspects of Memory, Swansea, Wales. Requests for reprints should be sent to Neal Kroll, Department of Psychology, University of California, Davis, CA 95616.

Of subjects who had chosen a foil suggested by MI, those who saw the sequential return between tests were more likely to correct their error on the second test than were those in the control conditions (.53 vs. .25 and .17). Thus, some subjects must retain OI after choosing a foil suggested by MI. To see whether the sequential return affected primarily true memory or merely biases, Kroll and Timourian (1986) used various assumptions to generate models. The best fit to the data included these assumptions: (1) MI does not affect memory, only bias; (2) OI memory deteriorates between tests for control groups; (3) bias diminishes over tests for all groups, indicating that postevent memory also deteriorates; and (4) order information given between tests provides an impressive retrieval cue, allowing true memory scores to increase on the second recognition test.

The coexistence hypothesis, defined by Loftus and Loftus (1980), includes not only the notion that both OI and MI continue to coexist, but also the idea that the major cause of errors is MI interference. However, the above results indicate that deterioration is due not to interference, but to bias. The return effect seems to be primarily an improvement of memory through a reinstatement of the original context—so that there are fewer trials on which to guess—and secondarily the result of an overall reduction in the biasing. Before these conclusions can be accepted, however, several weaknesses need to be addressed. In the present experiment, we (1) replicated with different slides to ensure that the results were not a peculiarity of the set used; (2) added a NI control (the model assumed that the degree of positive bias from CI was the same as the degree of negative bias from MI, but this is not always the case; e.g., Loftus et al., 1978); and (3) tested the parameters of the model over data sets with different overall guessing rates.

Because Kroll and Timourian (1986) found no difference between the two control procedures, a single control group was used here. This control group saw the return slides between tests, but in random order. All subjects rated their confidence in their recognition decisions.

METHOD

The subjects were 288 students in introductory classes who received extra credit for participating. They were tested in groups of 12 or fewer.

Experimental conditions, ignoring counterbalancing, formed a $2 \times 2 \times 3$ design. Subjects saw two or three alternatives (2A or 3A) during recognition tests, and saw the return slides in either sequential or random order. The third factor was within subjects: Subjects answered CI questions for three of the critical slides, MI for three, and NI for three.

Memory stimuli were 35 slides depicting a person returning home to find a crime in progress. There were nine critical slides, with three versions of each. These versions differed by one feature; for example, in one set a man in a window wore a ski cap, a baseball cap, or a fisherman's cap. Slides were counterbalanced so that subjects in each condition saw each version as the memory slide. Critical questions were each paired with a critical slide: three with CI, three with MI, and three with NI about the significant features. The paired slides are referred to as CI, MI, and NI slides. Each suggestion type was given for each critical slide equally often across subjects in each of the experimental conditions.

The 30 return slides were of the same scenes, but missing the critical information. For example, the return slide corresponding to the above example showed the same window, but without the man.

The test slides consisted of 15 slide sets, each consisting of three slides that differed in one feature. Nine of these test sets were of critical slides. Subjects in the 3A test conditions saw all three of each set; those in the 2A test conditions saw only two, the correct slide and a foil.

Subjects were told they would see slides of what one might see coming home. They were told to pay close attention; they would be ques-

tioned about what they had seen. After the event slides, they received a form with 20 questions. Each required a yes or no answer and a 1-5 confidence rating. After a 10-min distractor task, the subjects received the first recognition test, Test 1. On each trial, they chose which slide had been in the event slides and rated their confidence on a 1-5 scale. The test trials were in a different sequence from that of the memory slides; that is, they no longer followed the story line. For subjects in the 2A test condition, the foils for the three MI critical slides corresponded to the MI given in the questions.

The subjects were then told that the questions had included MI, which may have led to Test 1 errors. They were told that they would be retested after they "returned to the scene of the crime" via slides without critical information. Half the subjects in both the 2A and 3A groups saw return slides in random order; the other half saw them in sequence. On Test 2, they again chose the memory slide and rated their confidence. The Test 2 slides were in the Test 1 sequence, but arranged differently on the screen; that is, a slide on the far left during Test 1 would now be in the middle.

RESULTS

The probability of a correct choice on the recognition tests is given in Table 1 as a function of the suggestion made about the critical slide. The averages for Test 1 are based on all subjects who received the same number of alternatives on their recognition tests. For Test 2, subjects are divided by the type of return sequence (random or sequential).

Subjects scored higher than is usual in this type of experiment (scores are often as low as 70% in the NI condition). To demonstrate a MI effect on memory, McNemar tests were performed on Test 1. In 3A conditions, more subjects scored lower on MI than on CI slides (39 vs. 9, $p < .01$) or NI slides (37 vs. 19, $p < .02$). In 2A conditions, more subjects scored lower on MI than on CI slides (22 vs. 9, $p < .01$), but not significantly more than the number scoring lower on NI slides (19 vs. 10, $p \approx .10$). Thus, although subjects remembered a great deal, Test 1 performance was affected. The next question was whether the different return conditions resulted in differential effects on Test 2.

Chi-squares compared return conditions for the likelihood of subjects' being correct on Test 2 given an error on Test 1. To simplify statistical interpretation, only subjects with one Test 1 error of the relevant suggestion type were included in an analysis. Thus subjects with two Test 1 errors on NI slides were not included in the NI analysis. Those with one error on a NI slide and one on a MI slide were included in each of these separate analyses, but not in the combined analysis. This reduction made analyses considerably more conservative. All chi-squares have one degree of freedom and continuity correction. (Similar significance patterns are found with McNemar tests that compared subjects across return conditions.)

MI slides: Subjects in the 3A group with sequential return slides were more likely to correct their error on Test 2 than those with random return [$\chi^2 (N=57) = 6.711, p < .01$]. This is not significant in the 2A group [$\chi^2 (N=38) = 0.033, p > .10$], or with groups combined [$\chi^2 (N=95) = 3.546, p < .10$].

NI slides: Again sequential return subjects were more likely to correct a mistake on Test 2. As above, this difference is significant for 3A [$\chi^2 (N=41) = 6.065, p < .025$] but not for 2A [$\chi^2 (N=25) = 0.205$] subjects, and is significant for combined 2A and 3A groups [$\chi^2 (N=66) = 6.132, p < .025$].

MI and NI slides combined: The pattern is the same as that for NI slides: for 3A subjects, $\chi^2 (N=98) = 14.58, p < .001$; for 2A subjects, $\chi^2 (N=63) = 0.014, p > .10$; and for combined 2A and 3A groups, $\chi^2 (N=161) = 10.16, p < .005$.

Note that the improvement resulting from seeing the sequential return slides was as great for NI as for MI slides. This makes it more reasonable to assume that there was an overall retrieval improvement than to assume only a reduction in interference from MI in the incorrect suggestion.

Table 2 presents the estimated values for the probability of having a true memory (P_t), the probability of being correct by guessing, and guessing bias. These values are based on three assumptions: (1) Subjects are correct on NI slides either by remembering OI (a true memory) or by guessing correctly; that is, $P_c = P_t + (1/a)(1 - P_t)$, where P_c is the probability of being correct and a is the number of alternatives. Then, $P_t = (a * P_c - 1) / (a - 1)$. Thus, P_c was .905 for 3A subjects on Test 1 NI slides, giving $P_t = .858$; and the probability of being correct by chance is $.905 - .858 = .047$. (2) p_t is the same on MI slides, but the guessing probability is less because of a negative bias from those who do not remember OI, but do remember MI. Thus, 3A subjects had 86.8% correct choices on Test 1 with MI slides. Assuming that their P_t was .858 (estimated from NI slides), they had only a $.868 - .858 = .01$ chance of guessing correctly; that is, they had a $.01 - .047 = -.037$ guessing bias. (3) The guessing bias is positive with CI slides.

Although this model is not really tested, it is useful for providing a potential explanation of the underlying processes. For example, it suggests that P_t increases after sequential slides and decreases after random slides. With 3A groups, it also seems that the biasing effect decreases from Test 1 to Test 2.

Table 3 presents average confidence ratings for 3A subjects as a function of the choice correctness. These ratings are averaged over the different question types. It is likely that question type affects confidence, but the small number of errors for some suggestion types made further data breakdown impossible. The tabled values are averaged over all relevant subjects and slides. In the analyses, however, data were reduced: When comparing ratings for slides on which subjects were correct on both tests (C,C) with those ratings for slides on which they were correct on Test 1 but not on Test 2 (C,E), subjects were included in the analysis only if they had at least one slide with each of the two decision patterns, that is, C,C and C,E. A similar data reduction was made for each of the listed comparisons. This means that the average rating for the C,C pattern was different when compared with the C,E pattern than when compared with E,C and E,E patterns, but not by much.

For the C,C versus C,E comparison, there are 21 subjects in the random and 10 in the sequential group. Random sub-

Table 1
Percentage of Correct Choice as a Function of Suggestion,
Number of Alternatives, First Versus Second Test,
and Sequential Versus Random Return

Suggestion	Test 2		
	Test 1	Sequential	Random
Three Alternatives			
Correct Information	92.6	94.9	87.5
No Information	90.5	94.9	84.7
Misleading Information	86.8	94.0	82.9
Two Alternatives			
Correct Information	96.3	95.8	95.8
No Information	94.2	96.8	91.7
Misleading Information	91.2	92.1	89.8

Table 2
Estimations of Probabilities of True Memory, Correct Guessing,
and Guessing Bias Based Upon Model Presented in Text

Suggestion		Test 2		
		Test 1	Sequential	Random
Three Alternatives				
True	(All)	.858	.924	.771
Guess	NI	.047	.025	.076
Bias	CI	.021	0	.028
	MI	-.037	-.009	-.019
Two Alternatives				
True	(All)	.884	.935	.833
Guess	NI	.058	.032	.083
Bias	CI	.021	-.009	.042
	MI	-.030	-.046	-.019

Note—NI = no information. CI = correct information. MI = misleading information.

jects used higher confidence ratings than sequential subjects [$F(1,29) = 6.645, p < .025$], and the C,C pattern resulted in higher ratings than the C,E pattern [$F(1,29) = 33.885, p < .001$]. The significant interaction [$F(1,29) = 4.274, p < .05$] indicates that the difference between C,C and C,E patterns is greater for sequential than for random subjects. Thus, not only are sequential subjects more likely to be correct on Test 2 than are random subjects, they are also likely to be less certain of their errors. This is at least partially due to a differential selection factor across groups; that is, even on Test 1, before return slides, confidence is higher on C,E slides for random than for sequential subjects. This indicates that sequential subjects are unlikely to move from a correct decision on Test 1 to an error on Test 2 except for those Test 1 decisions on which they had very low confidence. Random return subjects, on the other hand, moved from correct decisions to errors even with a relatively high confidence on their first decision. In addition, sequential subjects decreased their confidence on C,E slides, whereas random subjects increased their confidence on these slides, but this three-way interaction was not significant [$F(1,29) = 1.942, p > .10$].

For the C,C versus E,C comparison, there are 7 subjects in the random and 20 subjects in the sequential group. Only the difference between decision patterns is significant, with C,C resulting in higher confidence ratings than E,C [$F(1,25) = 13.420, p < .01$].

For the C,C versus E,E comparison, there are 29 subjects in the random and 19 in the sequential group. The difference between decision patterns is significant, with C,C again resulting in higher ratings [$F(1,46) = 66.524, p < .001$]. Ratings are higher on Test 2 than on Test 1 [$F(1,46) = 9.401, p < .01$], and this increase in confidence is strongest with the E,E pattern [$F(1,46) = 6.141, p < .025$]. There is also a tendency [$F(1,46) = 2.883, p < .10$] for random subjects to be the more likely to show this pattern (i.e., random subjects had higher ratings on the second test of the E,E pattern). As with the C,C versus C,E comparison, this indicates that a sequential return not only increases the probability of being correct, but also improves the relationship between confidence rating and decision correctness.

DISCUSSION

Loftus's (1975) paradigm no longer seems the definitive demonstration of memory replacement that many had hoped it was. If an absence

Table 3
Average Confidence Ratings for Subjects Tested
with Three Alternatives

	First/Second Choices			Error/ Error
	Correct/ Correct	Correct/ Error	Error/ Correct	
Sequential Return Between Tests				
Test 1	4.78	3.14	3.45	2.86
Test 2	4.79	2.93	3.73	3.32
Gain (2-1)	.01	-.21	.27	.45
Random Return Between Tests				
Test 1	4.70	3.75	3.67	3.89
Test 2	4.78	4.14	4.13	4.76
Gain (2-1)	.08	.39	.46	.87

of order cues forces more subjects to guess (e.g., Bekerian & Bowers, 1983), then the biasing effect is afforded a large role, and the errors give an impression that the event is remembered as described by the postevent information (e.g., McCloskey & Zaragoza, 1985). Indeed, given the similarity of the target and foil, the changed sequencing, and the related postevent information about the critical detail, it may be that even the subject has the impression that a true memory has occurred.

If order information is provided *after* Test 1, scores improve and there is little difference among the scores on MI, CI, or NI slides (e.g., Kroll & Timourian, 1986). This is expected if a sequential return provides cues to improve OI retrieval, thereby reducing the number of guesses and, hence, the role of biasing. Biasing would probably play a major role even in the presence of order cues if p_r were lower than in the present experiment. One way to reduce p_r , allowing a stronger bias effect, is to lengthen the interval between event slides and questions, keeping the interval between MI and recognition test relatively short (E. F. Loftus et al., 1978).

Tulving (1983) explained the MI effect with his concept of *recoding*, defined as "all the processes and operations that take place after the encoding of the original event and that bring about changes in the engram" (p. 164). Tulving elaborated when discussing the MI effect:

Recoding of the engram implies that it is changed, and after the change has taken place some of the information in the original engram is no more available. . . . Utilization of certain information originally contained in it should not be possible after recoding has taken place. To prove the recoding hypothesis wrong, therefore, all that is necessary is to show that all originally stored information is still functionally intact after recoding has taken place. (Tulving, 1983, p. 168)

Hintzman (1984) considered recoding superfluous, preferring Semon's (1923) *nondifferentiating homophony*, in which similarities among engrams are emphasized by mutual reinforcement of common properties and mutual interference of their distinguishing ones. An advantage of homophony is that

it would allow the concept of recoding to be dropped. Retroactive interference, including . . . misleading questions in eyewitness-testimony experiments, may not be caused by recoding at all; it may simply reflect the simultaneous activation of the target and interpolated traces. If the information sought in the target trace conflicts with information in the interpolated trace, homophony-induced interference is likely to occur. (Hintzman, 1984, p. 241)

The present evidence argues against the hypothesis of recoding. If there is any effect here beyond that of a simple guessing bias—that is, if a subject has an impression of having a true memory—it is better explained by homophony than recoding. Indeed, the notion of homophony seems to better fit Tulving's (1983) model of "synergistic ephory," in which what is retrieved depends upon the joint effects of trace and retrieval information. If order information is distorted, OI is not readily available and postevent information plays a larger role—perhaps simply a changed guessing bias, perhaps homophony; the difference between "guessing" and "homophony" is defined in terms of a subject's having or not having a feeling of a true memory experience. If order information is reinstated, the p_r increases and the effect of postevent information diminishes.

The importance of context reinstatement on memory has been long known (e.g., McGeoch, 1932) and has been shown recently in situations with direct consequences for eyewitness memory (e.g., Davies & Milne, 1985). The effect of context on recognition is not as unambiguous as the effect on recall (e.g., Godden & Baddeley, 1980), but the role of order information in improving recognition memory has been shown both in the eyewitness experiment (e.g., Bekerian & Bowers, 1983) and in more traditional laboratory experiments (e.g., Jacoby & Hendricks, 1973). The importance of context cues for both recall and recognition should be expected if retrieval is required for both types of tasks (e.g., Tulving, 1983, p. 322); improving context cues should increase the likelihood of synergistic ephory.

The practical implications of our results are similar to those drawn by Davies and Milne (1985): Reinstating original context—here by sequential return slides—is an effective way of improving eyewitness memory. This will not, however, guarantee perfect memory. Some witnesses continue to choose an incorrect alternative with high confidence. The more theoretical implication is that Loftus and Loftus (1980) were premature in concluding that a preponderance of evidence now opposes the permanence of human memory.

REFERENCES

- BEKERIAN, D. A., & BOWERS, J. M. (1983). Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*, 139-145.
- BRIGHAM, J. C., & WOLFSKEIL, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. *Law & Human Behavior*, *7*, 337-349.
- COLE, W. G., & LOFTUS, E. F. (1979). Incorporating new information into memory. *American Journal of Psychology*, *92*, 413-425.
- DAVIES, G., & MILNE, A. (1985). Eyewitness composite production: A function of mental or physical reinstatement of context. *Justice & Behavior*, *12*, 209-220.
- GODDEN, D. R., & BADDELEY, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology*, *71*, 99-104.
- HINTZMAN, D. L. (1984). Episodic versus semantic memory: A distinction whose time has come—and gone? *Behavioral & Brain Sciences*, *7*, 240-241.
- JACOBY, L. L., & HENDRICKS, R. L. (1973). Recognition effects of study organization and test context. *Journal of Experimental Psychology*, *100*, 73-82.
- KROLL, N. E. A., & TIMOURIAN, D. A. (1986). Misleading questions and the retrieval of the irretrievable. *Bulletin of the Psychonomic Society*, *24*, 165-168.
- LOFTUS, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*, 560-572.
- LOFTUS, E. F., & LOFTUS, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist*, *35*, 409-420.
- LOFTUS, E. F., MILLER, D. G., & BURNS, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, *4*, 19-31.
- LOFTUS, E. F., & SCHOOLER, J. W. (1984). Recoding processes in memory. *Behavioral & Brain Sciences*, *7*, 246.
- 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*, 3-18.
- MCGEOCH, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, *39*, 352-370.
- SEMON, R. (1923). *Mnemonic psychology* (B. Duffy, Trans.). London: Allen & Unwin.
- STERN, W. (1904). Wirklichkeitsversuche. *Beiträge zur Psychologie der Aussage*, *2*, 1-31.
- TULVING, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.
- TULVING, E. (1984). Relations among components and processes of memory. *Behavioral & Brain Sciences*, *7*, 257-268.
- ZARAGOZA, M. S., MCCLOSKEY, M., & JAMIS, M. (1987). Misleading postevent information and recall of the original event: Further evidence against the memory impairment hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 36-44.