

Contextual overlap and eyewitness suggestibility

KAREN J. MITCHELL

Yale University, New Haven, Connecticut

and

MARIA S. ZARAGOZA

Kent State University, Kent, Ohio

Studies of eyewitness suggestibility have traditionally used a paradigm that maximizes the extent to which the postevent interview overlaps with the witnessed event in terms of narrative content, narrative structure, and environmental context. The present study explored whether these dimensions of overlap contribute to people's tendency to confuse suggested details for those they have actually witnessed. We systematically manipulated the extent to which the postevent questionnaire overlapped with the witnessed event. Across two experiments, overlap in narrative content, narrative structure, or environmental context was not found to increase suggestibility effects, even though the manipulation did have other memory effects (e.g., it improved cued recall of the actual source of the suggestions, Experiment 2). These findings suggest that understanding the interaction between the structure and content of the objective context in which misinformation is encountered and various remembering contexts (e.g., recognition vs. recall) is important for advancing our understanding of source confusion in an eyewitness situation.

Eyewitness memory is prone to contamination from other sources. For example, scores of studies have documented that exposure to postevent misinformation can lead people to misremember witnessing details and events that were merely suggested to them (see, e.g., Lindsay, 1994; Loftus, 1997; Zaragoza, Lane, Ackil, & Chambers, 1997, for reviews). It is now well accepted that the suggestibility of eyewitness memory reflects, in large part, the more general difficulty people have discriminating between related sources of information in memory (see, e.g., Belli & Loftus, 1994; Ceci, 1995; Johnson, Hashtroudi, & Lindsay, 1993; Mitchell & Johnson, 2000, for reviews). From the source monitoring perspective, people come to have illusory recollections of witnessing suggested events because they misattribute memories derived from the postevent source to the witnessed event.

Research and theory on source monitoring have elucidated many of the general mechanisms that underlie source confusion errors (see Johnson et al., 1993; Johnson & Raye, 2000; Mitchell & Johnson, 2000, for reviews). According to the source monitoring framework (Johnson et al., 1993), there is no single aspect of our memories that specifies its true source without fail. Rather, our memory records contain various characteristics or features that provide clues to their origin. People can, and often do, accurately attribute the source of their memories because memories from different sources tend to differ on average in the quantity and quality of the characteristics associated with them (e.g., memories of witnessed events typically have more vivid perceptual, temporal, and spatial information than memories of imagined events, e.g., Johnson, Foley, Suengas, & Raye, 1988). Nevertheless, because there can be overlap in the distributions of the features associated with memories from different sources, errors can occur. For example, imagining words spoken in another person's voice increases people's tendency to confuse what they imagined the person said with what they actually heard the person say, presumably because it increases the overlap between the characteristics of the two sources of information (Johnson, Foley, & Leach, 1988). Although feature overlap is not the sole determinant of source misattribution errors (e.g., retrieval conditions and judgment/decision processes also play a critical role in source monitoring accuracy; see, e.g., Dodson & Johnson, 1993; Marsh & Hicks, 1998; Zaragoza & Lane, 1998; see also Lindsay & Johnson, in press, for discussion of this point), there is extensive empirical evidence that points to feature overlap as a key factor in source confusion errors (e.g., Finke, Johnson, & Shyi, 1988; Hashtroudi,

Experiment 1 was included in the doctoral dissertation submitted by K.J.M. to Kent State University in partial fulfillment of doctoral requirements. Various portions of this project were supported by research awards from the Applied Psychology Center of Kent State University to K.J.M., NSF Grant SES-9819303 to M.S.Z., and NIA Grant AG09253 to Marcia K. Johnson. Special thanks to Marcia Johnson and Carol Raye for inspirational conversation throughout the writing process; we thank the members of MEMlab2000 for helpful comments on an earlier draft. The assistance provided by the Psychology Departments of Mercyhurst College and Penn State Erie via use of their subject pools is gratefully acknowledged. Thanks to Sarah Drivdahl, Kristie Payment, and Erin Noonan for Experiment 2 data collection and Molly Maxfield for Experiment 2 data entry. Correspondence should be addressed to K. J. Mitchell, Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205 (e-mail: karen.mitchell@yale.edu) or M.S. Zaragoza, Department of Psychology, Kent State University, Kent, OH, 04220 (e-mail: mzaragoz@kent.edu).

Johnson, & Chrosniak, 1990; Henkel & Franklin, 1998; Johnson, Foley, & Leach, 1988; Koutstaal & Schacter, 1997; Lindsay, Johnson, & Kwon, 1991; Markham & Hynes, 1993; Rabinowitz, 1989; Winograd, 1968; see also Johnson et al., 1993; Johnson & Raye, 2000; Mitchell & Johnson, 2000, for reviews).

In studies of eyewitness suggestibility, the featural overlap between the content and context of the two sources (i.e., the original event and postevent interview) is multifaceted and extensive. For example, the two episodes are intimately related because they share a common referent—the eyewitness event. Note that the common referent factor is inherent in every eyewitness interrogation because, by definition, the postevent interview is always *about* the witnessed event.

A related point is that there is a great deal of overlap in the narrative structure and content of the two episodes. Specifically, in the typical study participants read through a narrative or questionnaire that describes each of the scenes of the video in chronological order and in great detail, with the misleading details interspersed throughout this narrative account. Hence, the information they encounter in the postevent interview overlaps extensively in structure and content with that derived from the witnessed event. Although some narrative overlap is to be expected in any postevent interview situation, the procedures used in laboratory studies of eyewitness suggestibility have tended to maximize the narrative similarity of the two episodes.

Finally, in the typical eyewitness suggestibility study, the witnessed event and postevent questioning occur close together in time, in the same room, with the same experimenter, same co-participants, and so on. This high level of environmental overlap between the two sources may also contribute to the confusability of the two episodes.

In summary, in the eyewitness suggestibility paradigm there are multiple sources of featural overlap between the originally witnessed event and the postevent interview, any (or all) of which may contribute to the confusability of the two sources. However, almost nothing is known regarding the amounts of these various types of overlap that are required for misattribution errors to occur. Yet such information is critical in developing a theoretical understanding of eyewitness suggestibility and in generalizing from laboratory studies to real-world situations. Thus, the objective of the present study was to begin to evaluate the contribution of various sources of overlap to these errors. To this end, in Experiment 1 the extent to which the narrative structure and content of the postevent interview overlapped with that of the witnessed event was systematically varied. We chose to manipulate these factors because they are the types of overlap that can be expected to vary across interview situations in the real world. For example, because the goal of most forensic interviews is to gather information so that the “story” can be put together, it seems unlikely that an interrogator would provide misinformation embedded in the kind of

complete and richly detailed narrative account of the witnessed event typically used in laboratory studies.

EXPERIMENT 1A

Participants viewed a video depicting a house burglary and subsequently answered a postevent questionnaire containing misleading suggestions, some of which were repeated (see Zaragoza & Mitchell, 1996, for evidence that repeated suggestion increases suggestibility). The key manipulation involved the nature of the postevent questions. Across four groups, the specific suggestions used were exactly the same; what varied was the amount of narrative overlap between the video and the postevent questions. At one extreme (high-narrative-overlap condition) participants received a questionnaire that provided a detailed and coherent retelling of the original event that, with the exception of the few pieces of presupposed misinformation, was a rich and veridical account of the events that participants actually witnessed. At the other extreme (low-narrative-overlap condition) participants were questioned about the events of the video in random, rather than chronological, order, and most of the narrative text was deleted. The other groups were presented with intermediate levels of narrative overlap, as described below. If narrative overlap contributes to suggestibility in this paradigm, source misattribution errors should decrease when narrative overlap is reduced.

Method

Participants and Design

Two hundred eighty-one undergraduates were randomly assigned to one of four narrative overlap groups, as described below. Amount of narrative overlap thus served as a between-participants factor. Number of exposures to misleading suggestion was manipulated within participants; for any 1 participant, four suggestions appeared once, four appeared three times, four other critical items served as never-presented control items. This resulted in a 4 (overlap group) \times 3 (number of exposures) mixed design.

Materials and Procedure

Participants were run in small groups (≤ 10) with participants from each of the overlap groups intermixed.

Phase 1—The eyewitness event. The eyewitness event was a 5-min segment of a police training video (Zaragoza & Mitchell, 1996). It depicts a burglary of a home by two youths and an ensuing police car chase.

Phase 2—Misleading postevent questioning. Immediately after seeing the video, participants answered 36 printed questions that varied by group, as described below. For all groups, the questionnaire was composed of three 12-question subsets. Each question in a subset referred to 1 of 12 unique events in the video (e.g., the thief leaving the home) and corresponded to 1 of 12 misleading postevent suggestions (the thief wore gloves, the thief pulled a window shade, the thief stole a ring, the thief had a gun, the driver smoked a cigarette, there was a barking dog, the thief put on his seatbelt, the neighbor’s name was Mrs. Anderson, one of the police officers was drinking coffee, the police said they would shoot, the driver was DUI, the driver jumped a curb with the car).

One subset followed the other without interruption. Thus, the questionnaire essentially reviewed the original events for the participants three times in succession. All participants answered all 36 questions,

but by adding suggested items to specific questions as necessary, number of exposures to these items was manipulated within participants. Hence, each participant received only a subset of the 12 suggested items. That is, for each participant, four suggested items were presented in all three subsets of questions (three-exposure level), four were presented in only the last subset of questions (one-exposure level),¹ and finally, four served as never-presented control items. Counterbalancing assured that all items served at all exposure levels equally often across participants.

High narrative overlap (hereafter referred to as “high overlap”; $n = 72$). This group received the same 36-item questionnaire used by Zaragoza and Mitchell (1996). Each question contained a substantial amount of text that overlapped with the narrative content of the video (e.g., “Later, as he was leaving the house the thief, putting his hand on the *gun* at his waist, looked both ways and went out the door. Did he slam the door behind himself?”; see Appendix for other examples). Note that in the video the thief really was shown leaving the house, looking both ways, going out the back door, and slamming the door behind him; that is, with the exception of the suggestion about the gun, the details in this question are true in terms of the events in the video. Note too that the order of these questions retained the chronology of the events in the video, and many explicit temporal cues were provided (e.g., “At the beginning of the scene . . .,” “Next . . .,” “Meanwhile . . .”), all of which should have helped participants “relive” the events of the video quite easily. Thus, the high-overlap questions essentially reinstated the original event for the participants in elaborate detail.

Moderate narrative overlap (hereafter referred to as “moderate overlap”; $n = 72$). The purpose of this condition was to reduce the extent to which the postevent review matched the narrative content and structure of the video. To this end, we disrupted the cohesiveness of the story line by removing the temporal cues that served as links between scenes (e.g., “Meanwhile,” “At the same time,” “Then,”). We also reduced the number of true narrative details (e.g., “The thief with the *gun* went out. Did he slam the door?”; see Appendix for other examples). There was a constraint in that we had to ensure not only that we were suggesting the same items to all participants, but also that the manipulation of narrative overlap did not disrupt the nature of the suggestions. Specifically, the suggestions had to be (1) presupposed in the prefatory text (see, e.g., Loftus, 1981, for evidence that the level of suggestibility is affected by whether the suggestion is the focus of the question or presented as a supposition) and (2) tied to the same video scene as in the high-narrative-overlap questions. Note, then, that although this manipulation served to disrupt the cohesiveness of the narrative, it is best thought of as a manipulation that reduced, but did not completely eliminate, narrative overlap with the video.

Moderate overlap plus irrelevant information (hereafter referred to as “moderate + irrelevant”; $n = 65$). One incidental consequence of reducing narrative overlap in the manner just described was that the moderate-overlap questions contained fewer words than did the high-overlap questions. The intent of the moderate + irrelevant group was to control for the number of words in the postevent questions to ensure that any reduction in errors in the moderate-overlap group was not merely the result of participants’ being exposed to fewer verbal material (e.g., by decreasing interference or increasing the salience of the suggested information; see, e.g., Tournant, Hall, & Loftus, 1986). That is, better source memory performance could obtain in the moderate-overlap condition for reasons not directly related to the amount of narrative overlap, per se.

Therefore, this group received a questionnaire that was identical to that used for the moderate-overlap group, except that each question was accompanied by enough extra instructional information to bring the total number of words up to that of the corresponding item on the high-overlap questionnaire (e.g., “Do not skip any questions

and do not provide ‘I don’t know’ as an answer. The thief with the *gun* went out. Did he slam the door?”; see the Appendix for other examples). To ensure that all participants across groups got the question last, this instructional information was always presented before the prefatory text in which the suggested item was embedded. The extraneous text was indistinguishable in terms of font, position on the page, and so on. It is interesting to note that whereas the postevent questions and video were dissimilar in the other groups by virtue of what was missing from the questions, the video and questions were dissimilar in this group because the questionnaire contained a substantial amount of information that was irrelevant to the witnessed event. If it is the nature of the information in the questions (i.e., its narrative overlap with the video) that is important, as opposed to merely the amount of information, suggestibility effects should be attenuated in this group also.

Low narrative overlap (hereafter referred to as “low overlap”; $n = 72$). One might argue that because the moderate-overlap questions retained the general chronology of the video and some of the content, there was still substantial narrative overlap between the two events. Thus, to further reduce narrative overlap with the video in the low-overlap group, the moderate-overlap questions were randomly ordered within each subset (see the Appendix for examples). This served to severely reduce the cohesiveness of the story line, making it much less similar to the narrative structure of the witnessed event. There was a constraint that no two successive questions referred to the same event or contained a suggested item of the same exposure level.

Participants completed the questionnaire at their own pace. They were cautioned that they should be sure to read all of the information before answering and that they should answer each and every question even if they had to guess.

Phase 3—Source memory test. A surprise source recognition memory test followed a 10-min filled interval. The test was composed of an equal number of items from each of the four possible source categories (i.e., video only, questions only, both, neither). The probes were 32 pairs of statements read on a cassette recorder in a male voice. Both items of a pair referred to the same information; one probe of a pair asked about the participants’ memory for the video, and the second asked about their memory for the questions. The order never varied (i.e., “In Video?” followed by “In Questions?”). Twelve pairs of statements contained the critical items. For example, participants would hear, “Number 3. In the video, the thief had a gun,” and 4 sec later they would hear, “Number 3. The questions said the thief had a gun.” For any 1 participant, four of these critical pairs contained control items (0 exposures) and eight contained suggested items (four each at 1 and 3 exposures). There were also 20 pairs of filler items: 8 video only, 8 from both the video and questions, and 4 new items. Note that the four novel items, together with the four suggested items not exposed, comprised eight *new* (or *neither* source) items for each participant. The interitem interval was 4 sec. Probes were presented in the same random order to everyone. The constraints on randomization were that a suggested item could not occupy the first two or last two positions, and no two consecutive pairs of probes could be from the same source category.

Participants were given both written and verbal instructions for the source memory test. They were explicitly informed that some of the statements they would hear contained information that was only in the video, some contained information that was not in the video but was contained in the questions they answered, some contained information that was in both the video and the questions, and finally, that some of the test statements contained information that was in neither the video nor the questions. Furthermore, participants were told that they were to base their source judgments on their own memory of the events. Including such an explicit warn-

Table 1
Mean Proportion of Suggestions Attributed to the Video
(i.e., Source Misattributions Collapsed Across Confidence)
by Condition and Number of Exposures for Experiment 1

Condition	No. Exposures		
	0	1	3
Experiment 1A			
High	.09	.24	.42
Moderate	.07	.29	.42
Moderate + irrelevant	.09	.27	.42
Low	.09	.27	.38
Experiment 1B			
High	.22	.49	.65
Low	.19	.46	.57

ing about misinformation increases our confidence that we are measuring false memory and not some other aspect of suggestibility, such as demand.

Participants responded on an answer sheet that contained two columns, clearly labeled “video” and “questions.” Each column contained 7-option Likert-type scales (*definitely yes, probably yes, maybe yes, unsure, maybe no, probably no, definitely no*).

Our interest was in the pattern of participants’ memory for the source of the suggested items as a function of overlap group and exposure level. For this reason, only participants’ memory for the suggested items will be reported. A “yes” response to a suggested item in the Video column signals a source misattribution error, and a “yes” response to an exposed suggestion in the Questions column indicates a correct source judgment. The responses are not mutually exclusive; one can recognize the correct source of an item and still misattribute it.

Results and Discussion

To foreshadow, the hypothesis that amount of narrative overlap influences the source confusion seen in suggestibility studies was not supported. Reducing narrative overlap between the original event and the postevent episode by reducing the number of true content details and cohesiveness of the story, disrupting the chronology of events, or adding extraneous information had no discernible effect on participants’ suggestibility, even with repeated exposure. Indeed, reductions in narrative overlap had no effect on any of the dependent measures examined in this experiment. We report the data pooled across confidence levels (i.e., the sum of *maybe yes, probably yes, and definitely yes* responses); analysis of the data broken down by confidence level yielded an identical pattern of results. An alpha level of .05 was used for all analyses in this paper.

Source Misattributions

Misattributions of suggestions to the video were submitted to a 4 (overlap group: high, moderate, moderate + irrelevant, low) \times 3 (number of exposures: 0, 1, 3) mixed analysis of variance (ANOVA). The data are presented in Table 1. Consistent with past research (e.g., Mitchell & Zaragoza, 1996; Zaragoza & Mitchell, 1996; Zaragoza, Mitchell, & Drivdahl, 1997), there was a reliable main effect of number of exposures, indicating that the more often people were exposed to the suggestions, the more

likely they were to misremember seeing them in the video [$F(2,554) = 157.13, MS_e = .05$]. However, neither the main effect of group nor the group \times number of exposures interaction was reliable ($F_s < 1$). Thus, the expectation that suggestibility effects would differ as a function of amount of narrative overlap was not supported.

Memory for Actual Source

We also assessed whether reductions in narrative overlap influenced participants’ ability to accurately identify the questionnaire as the true source of the suggested items (measured as any “yes” response to suggested items in the Questions column). The data appear in Table 2. Once again, neither the main effect of group nor the group \times exposures interaction was reliable ($ps > .05$). As would be expected, participants were more likely to remember reading the suggested items in the questions the more often they read them [$F(2,554) = 2,886.59, MS_e = .02$].

Item Recognition

Finally, we examined whether there were group differences in participants’ ability to recognize the suggestions as having occurred in the experiment. Item recognition was defined as any “yes” response to a suggested item in either column; that is, a “yes” response to a suggestion in the Video column and/or the Questions column. As would be expected on the basis of the previously reported source memory data, item recognition was nearly identical across the four groups; neither the main effect of overlap group nor the group \times number of exposures interaction was significant ($ps > .05$). Once again, the only reliable effect was due to repetition: Participants were more likely to recognize suggestions the more often they had read them [$M_s = .10, .84, .96$, for the 0-, 1-, and 3-exposure levels, respectively; $F(2,554) = 2,479.41, MS_e = .03$].

EXPERIMENT 1B

In Experiment 1B, we explored the possibility that our manipulation of narrative overlap was not effective in Experiment 1A because the postevent narrative occurred immediately after participants had viewed the video. Perhaps because participants had just seen the video, people

Table 2
Mean Proportion of Suggestions Attributed to the Questions
by Condition and Number of Exposures for Experiment 1

Condition	No. Exposures		
	0	1	3
Experiment 1A			
High	.02	.76	.91
Moderate	.02	.81	.96
Moderate + irrelevant	.02	.77	.95
Low	.01	.84	.97
Experiment 1B			
High	.03	.69	.86
Low	.01	.72	.89

in the moderate- and low-overlap groups were able to reflectively “fill in” the missing narrative information with the same sort of information that was explicitly provided in the high-overlap group’s questionnaire. If participants in the groups with reduced narrative overlap were able to call up, spontaneously, detailed knowledge of the content and structure of the queried scene from the sparse cues provided in their questionnaire, the narrative overlap between the original event and postevent interview episodes might have been functionally similar across groups. Thus, in Experiment 1B we attempted to boost the effects of reducing narrative content in the questions by inserting a 24-h delay between the video and the postevent questions. We reasoned that because memory for the video could be expected to decay with time, participants would come to rely more on the external reinstatement of the details of the video when answering the questions after a 24-h delay, making the reduction of narrative overlap more powerful.

Furthermore, in Experiment 1A there was substantial environmental overlap between the original event and the postevent, given that both episodes took place during the same experimental session, in the same room, with the same experimenter, and so on. Thus, to reduce objective overlap even further, we added an environmental context shift between the two episodes for the low-overlap group. Although previous work by Bonto and Payne (1991) failed to yield support for the idea that reducing the overlap in environmental context alone would increase the discriminability of memories from the two sources, it is possible that such a context shift could be additive with the narrative overlap manipulation, leading to an attenuation of errors in the low-overlap group.

Method

Participants and Design

One hundred sixty-two undergraduates were randomly assigned to either the high-overlap or low-overlap group ($n_s = 81$) as described below. Amount of objective overlap thus served as a between-participants factor with number of exposures to misleading suggestion manipulated within participants. This resulted in a 2 (overlap group) \times 3 (number of exposures) mixed design.

Materials and Procedure

All participants underwent essentially the same three-phase procedure used in Experiment 1A, except that they were dismissed after viewing the video and returned the next day to answer the postevent questions. Overlap was manipulated between two groups as follows.

High overlap. This group received the same postevent questionnaire as the high-overlap group in Experiment 1A. Remember that it contained a substantial amount of true narrative information from the video and the events were queried in the same chronological order. When this group of participants returned for the misleading postevent questions on Day 2, they returned to the same room in which they had watched the video, they were tested with the same co-participants, and the same experimenter ran the session. The objective overlap for this group was thus maximized.

Low overlap. This group received the low-overlap questionnaire from Experiment 1A. Remember that this questionnaire drastically reduced the narrative overlap with the video by reducing the amount of true content information, removing temporal cues, and querying the events in random order. When this group of partici-

pants returned on Day 2, they went to a room different from that in which they had watched the video, they were tested with additional co-participants in a setting that was substantially different from that on Day 1 (e.g., a large lecture hall on Day 2 vs. a small classroom on Day 1), and a different experimenter ran the session (one who had not been present on the 1st day). The objective overlap for this group was thus minimized beyond that provided by the reduction of narrative overlap.

For both groups, the source test followed a 10-min filled interval.

Results and Discussion

The data of primary importance were, again, participants’ source misattribution errors. In short, we replicated the findings of Experiment 1A (Table 1). Although adding a retention interval between the original event and postevent questioning increased the overall incidence of misattribution errors in both groups of Experiment 1B (probably because of less resistance to the suggestions given poorer memory for the original event), the two groups did not differ reliably from each other, nor was the group \times exposures interaction significant ($p_s > .10$). Thus, once again, source misattribution errors were not attenuated when objective overlap was reduced. Only a reliable main effect of exposure level obtained [$F(2,320) = 109.54$, $MS_e = .06$]. We should note that although the present study provided only a limited test of the environmental context manipulation, the results converge with those of Bonto and Payne (1991) in showing no effect of a shift in environmental context.

There were also no reliable group differences, or significant interactions, for measures of memory for the actual source of the suggestions or item recognition (all $p_s > .10$). As with source errors, only a reliable main effect of exposure level obtained for memory for actual source [$F(2,320) = 1,038.36$, $MS_e = .03$; Table 2] and item recognition [$M_s = .22, .83, .94$, for the 0-, 1-, and 3-exposure levels, respectively; $F(2,320) = 677.36$, $MS_e = .04$].

In summary, the results of Experiments 1A and 1B collectively show that false memory for suggested (and repeatedly suggested) events was unaffected by substantial changes in the content, structure, and environmental context of the postevent interview. The pattern of results was remarkably consistent across all of the groups tested in these experiments, including the effect of repeated exposure to suggestion.

EXPERIMENT 2

In Experiments 1A and 1B, the measure of suggestibility was participants’ misattributions of the suggested items to the video on a source recognition test. In Experiment 2, we assessed whether narrative overlap would influence participants’ tendency to *recall* the suggested items when queried about their memory for the witnessed event 2 days after the postevent interview. The rationale for using a delayed cued recall measure of suggestibility was the possibility that the immediate source recognition test employed in the foregoing experiments was insensitive to group differences in confusion regarding the source

Table 3
Mean Proportion of Suggestions Recalled as Seen in
the Video and Read in the Questions in Experiment 2

	Video			
	Total Errors		Video Only	
	0 exposure	1 exposure	0 exposure	1 exposure
High	.07	.23	.07	.13
Low	.12	.33	.12	.20
	Questions			
	Total		Question Only	
	0 exposure	1 exposure	0 exposure	1 exposure
High	.02	.60	.01	.45
Low	.01	.47	.01	.26

of the suggested items. With very short retention intervals between the postevent misleading questions and the memory test, the familiarity of the suggested items was likely very high in all of the groups, especially when the suggestions were repeated. An overreliance on familiarity as a basis for attributing items to the video could have masked potential group differences in other features of memory for the suggested items. If the narrative overlap manipulation employed in Experiments 1A and 1B does influence the confusability of the source of the suggestions, it should be easier to detect group differences on a delayed recall test, where the retrieval demands are much greater.

Method

Participants and Design

One hundred thirty-four undergraduates were randomly assigned to either the high-overlap or low-overlap group ($n = 67$) as described below. Amount of overlap thus served as a between-participants factor. There were only two levels of exposure in this experiment: 0 exposures (control items) and 1 exposure (suggestions), manipulated within participants. This resulted in a 2 (high vs. low overlap) \times 2 (0- vs. 1-exposure) mixed design.

Materials and Procedure

Participants underwent the same basic three-phase procedure as did the high-overlap and low-overlap groups in Experiment 1A with the following exceptions:

After watching the videotape on Day 1, participants did an unrelated filler task for 5 min and then completed 12 postevent questions. The questions were one subset of the high- and low-overlap questionnaires used in Experiment 1A; they contained 10 critical questions (5 contained a suggestion and 5 were 0-exposure control items) and 2 fillers (i.e., "the neighbor's name was Mrs. Anderson" and "the driver was DUI" were not used as suggestions). Items served equally often in the high- and low-overlap conditions as suggested and control items. To equate exposure time/encoding across groups, participants were paced through the questions by a tape-recorded voice that read the questions with a 5-sec interval between questions. Participants used a paper to cover their questionnaires and revealed only one question at a time as the tape instructed. After they had answered all 12 questions they were dismissed. They returned 2 days later to complete the source memory test.

Upon their return on Day 2, participants were given a printed cued recall test that queried them about their memory of the video and the questionnaire. As in the source recognition test instructions,

the recall test instructions emphasized that some of the items were in only the questions, some were in only the video, some were in both, and some items were completely new. For each of 13 items (10 critical target items, 3 fillers), participants were asked about both their memory for the video and for the questions (in that order). For each item, participants were first asked whether they remembered, in the video, the general category of information associated with the target item. For example, for the suggestion that the thief had a gun, participants were asked: "In the video do you remember seeing the thief carrying a weapon?"; they responded by circling "yes," "no," or "not sure." For those items given a "yes" response, participants were told to write in the exact item(s) they recalled (e.g., "If yes, what weapon did you see? _____"). Immediately following, they were asked the same question about their memory of the questionnaire (e.g., "In the questions do you remember reading that the thief carried a weapon? ["yes," "no," or "not sure"]. If yes, what weapon did you read about? _____"). Participants completed the questions at their own pace.

Results and Discussion

For both memory for the video and memory for the questions, we extracted two measures for the 10 critical items: assent to the initial query about the general category of information representing the critical items (e.g., saying "yes" to the initial question asking whether the participant saw/read about the thief carrying a weapon) and proportion of initial assents that were followed by recall of a suggestion (e.g., writing in *gun* as the weapon seen/read about). We analyzed these data using 2 (high vs. low overlap) \times 2 (0 vs. 1 exposure) mixed ANOVAs.

Misattributions to the Video

With regard to the initial queries about the video, 3 participants (2 high- and 1 low-overlap) did not assent to any general queries associated with the suggestions. There were no group differences in the tendency to say "yes" to the initial queries about the suggested (i.e., 1-exposure) items ($M = .50$ and $.51$ for high- and low-groups, respectively).

Of primary interest was participants' tendency to recall having seen the suggested items in the video. For this measure, we computed for each participant the proportion of assents to the video query that were followed by recall of the suggested item. Participants who did not assent to any critical items were assigned a zero for recall. Contrary to the hypothesis that narrative overlap increases source misattributions, the high-overlap participants were *less* likely than were low-overlap participants to recall seeing the critical items in the video, as evidenced by a main effect of condition [$F(1, 132) = 5.62, MS_e = .07$; see Total Errors column in Table 3]. As expected, there was also a reliable suggestibility effect in that both groups were more likely to recall seeing the critical items in the video if they had been exposed to them in the questions than if they had not [$F(1, 132) = 25.87, MS_e = .09$; Table 3]. However, the interaction of overlap group and exposure level was not reliable ($F < 1$), showing that the manipulation affected misattributions to both the exposed (1-

exposure) and nonexposed (0-exposure) items equally, rather than simply reducing misattributions of the suggested items.

Note that in our procedure, one can incorrectly recall the suggestions from the video but also accurately recall reading the suggested items in the questions, and in many cases participants did so. For this reason, we also assessed participants' tendency to recall the critical items as being in the video but not in the questions (the only case in which recall was completely inaccurate) and found the same pattern: a reliable main effect of group [$F(1,132) = 4.88$, $MS_e = .05$] and exposure level [$F(1,132) = 5.12$, $MS_e = .07$], but no interaction ($F < 1$; see Video Only column in Table 3).

Memory for the Questions

Sixteen high- and 23 low-overlap participants did not assent to any general queries related to memory for having read about the suggestions in the questions. Group differences in initial assents to the general queries ($M_s = .28$ and $.21$ for the high- and low-overlap groups, respectively) were marginally reliable [$t(132) = 1.93$, $p = .06$].

With regard to recalling the suggested items from the questions, the pattern of group differences is a bit more complex. Overall, there was no difference between the groups in the proportion of their initial assents that were followed by recall of reading the suggestions, nor did group reliably interact with number of exposures ($p_s \geq .10$; see Total column in Table 3). Only the main effect of exposures was significant [$F(1,132) = 169.30$, $MS_e = .11$]. Note, however, that in many cases where participants accurately recalled reading the suggested item in the questions, they also incorrectly recalled seeing the same suggested item in the video. If one examines those cases where participants were completely accurate (they recalled the suggestions as being in the questions but did not [falsely] recall them from the video; see Questions Only column in Table 3), participants in the high-overlap group were significantly more accurate than were participants in the low-overlap group. In addition to a main effect of group [$F(1,132) = 6.65$, $MS_e = .09$], there was a reliable group \times exposure level interaction [$F(1,132) = 6.67$, $MS_e = .09$] that resulted because the high-overlap group was reliably more accurate than was the low-overlap group in recalling that the suggestions were read only in the questions [i.e., the 1-exposure level; $t(132) = 2.61$], even though the groups did not differ in the base rate of attributions to the questions. Finally, as expected, both groups were more likely to recall only reading the critical items in the questions if they had been exposed to them in the questions than if they had not [$F(1,132) = 88.41$, $MS_e = .09$].

In summary, with a delayed cued recall test, there was clear evidence that the narrative overlap manipulation employed in the foregoing experiments affected participants' memory for the source of the suggested items, but not in the expected direction. There was once again no evidence that high narrative overlap increases suggestibility. Rather, the high narrative overlap manipulation re-

duced source misattributions and improved participants' ability to remember that the suggestions had been encountered in the questionnaire only.

GENERAL DISCUSSION

Studies of eyewitness suggestibility have typically used a laboratory paradigm that maximizes the extent to which the postevent interview overlaps with the witnessed event in terms of its narrative content, narrative structure, and environmental context. Because it is well established that feature overlap contributes to source confusion errors more generally (e.g., Finke et al., 1988; Hashtroudi et al., 1990; Henkel & Franklin, 1998; Johnson, Foley, & Leach, 1988; Koutstaal & Schacter, 1997; Lindsay et al., 1991; Rabinowitz, 1989; Winograd, 1968; see also Johnson et al., 1993; Johnson & Raye, 2000; Mitchell & Johnson, 2000, for reviews), in the present study we explored whether these particular dimensions of overlap contribute to people's tendency to confuse suggested details for those they witnessed. To this end, we systematically manipulated the extent to which the postevent questionnaire overlapped with the witnessed event. Across several experiments there was no evidence that high amounts of overlap in narrative content, narrative structure, or environmental context produced larger suggestibility effects than low amounts of such overlap. In fact, across all of these experiments, the only reliable effect of narrative overlap (Experiment 2) was in the opposite direction, with participants in the high-overlap group making fewer source misattribution errors (and evincing more completely accurate source identifications) on a test of delayed cued recall than participants in the low-overlap group.

Given that increases in source overlap typically reduce source discrimination accuracy (see, e.g., Johnson et al., 1993; Johnson & Raye, 2000; Mitchell & Johnson, 2000, for reviews), it is interesting to consider possible reasons why the present studies in some cases failed to replicate this finding (Experiment 1) and in others produced the opposite result (Experiment 2). One possibility is that, at least in this situation, the relationship between source overlap and source confusion is not linear; rather, it may be better described by a step function. By this account, once the overlap between the two sources reaches a certain threshold, further increases in the amount of overlap do not increase confusion appreciably. Although the present experiments manipulated the extent to which the postevent questionnaires overlapped with the witnessed event, there were nevertheless several dimensions of overlap between these sources that were preserved across all groups. For example, in every group there was some overlap in the content of the original event and postevent questionnaire, because every questionnaire presupposed that the suggested item/event occurred in a specific scene from the video (e.g., instead of simply suggesting that the thief had a gun, in every group it was suggested that the thief had a gun in the scene where he left the house). Thus, the same individual scenes from the video were mentioned in every

group's questionnaire (albeit sometimes in rather sparse detail and in random order), providing some level of overlap in content.

Another possible source of overlap that remained constant across groups was that the postevent questionnaire was always about the witnessed event. Thus, the two sources of information were highly related because they shared a common referent (i.e., the witnessed event). Past research has documented that the semantic relatedness of two sources of information contributes to their confusability (e.g., Lindsay et al., 1991). Eyewitness interrogations lie at one extreme of the semantic relatedness continuum because in these situations both sources contain information about the very same event, not just information about the same, or related, topics. It may be that the high level of relatedness that results from sharing a common referent is sufficient to produce high levels of source confusion.

A related, but theoretically distinct, source of overlap may have been provided by the participants. In attempting to answer specific questions about each scene, participants undoubtedly reflected back on the original event (see Zaragoza & Lane, 1994). Under these circumstances, the low-overlap groups may have retrieved details of the scenes, mentally "filling in" information not provided by the low-overlap questionnaire. Thus, it may be that although, objectively, the overlap in narrative content varied substantially across groups, subjectively the content of the questioning episodes did not vary much across groups at all.

In summary, one potential explanation for the present findings is that the aforementioned remaining dimensions of overlap produced sufficient levels of confusability between episodes that additional overlap on other dimensions (narrative content and structure) did not reliably increase confusability. Understanding the relative contributions of these other dimensions of overlap to source confusion in an eyewitness situation remains an important issue for future research.

Another likely explanation for our seemingly discrepant findings is that the narrative *coherence* of the high-overlap questionnaire had memory-enhancing properties of its own. The results of Experiment 2 showed that, relative to participants in the low-overlap group, participants in the high-overlap group were better able to remember that suggested items had been encountered in the questionnaire only. Hence, any increases in source confusability that may have resulted from high narrative overlap were probably offset by improved memory for the suggestions' true source. A more direct test of the role of narrative overlap in suggestibility might be to compare high- and low-overlap groups under conditions where the questionnaires are matched in coherence. It seems likely that under such conditions high narrative overlap would lead to greater suggestibility than low narrative overlap.

With regard to coherence, it is interesting that the effects of narrative overlap varied depending on the type of test. Manipulations of narrative overlap produced no effects on a source recognition test, yet the same manipulation

yielded clear effects on a delayed cued recall test, with high overlap producing better source discrimination than low overlap. One intriguing possibility is that, whereas encountering the suggested information embedded in a highly embellished, organized, and coherent account of the witnessed event (as opposed to a more sparse and disorganized context) may lead to better source discrimination on tests that cue witnesses to recall specific details, it might be expected to increase the likelihood that people would spontaneously report the suggested items when freely providing narrative accounts of the events they witnessed. That is, the narrative coherence that led to improved performance on tests of cued recall might actually produce greater suggestibility when participants are tested with memory measures that are more heavily influenced by organizational and relational processes, such as free recall. In any event, the present findings suggest that understanding the potential interaction between the structure and content of the objective context in which misinformation is encountered and various remembering contexts (e.g., source recognition vs. cued recall vs. free recall) is a crucial next step for advancing our practical and theoretical understanding of source confusion in an eyewitness situation.

REFERENCES

- BELLI, R. F., & LOFTUS, E. F. (1994). Recovered memories of childhood abuse: A source monitoring perspective. In S. J. Lynn & J. W. Rhue (Eds.), *Dissociation: Clinical and theoretical perspectives* (pp. 415-433). New York: Guilford.
- BONTO, M. A., & PAYNE, D. G. (1991). Role of environmental context in eyewitness memory. *American Journal of Psychology*, **104**, 117-134.
- CECI, S. J. (1995). False beliefs: Some developmental and clinical considerations. In D. L. Schacter (Ed.), *Memory distortion: How minds, brains, and societies reconstruct the past* (pp. 91-125). Cambridge, MA: Harvard University Press.
- DODSON, C. S., & JOHNSON, M. K. (1993). Rate of false source attributions depends on how questions are asked. *American Journal of Psychology*, **106**, 541-557.
- FINKE, R. A., JOHNSON, M. K., & SHYI, G. C.-W. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory & Cognition*, **16**, 133-137.
- HASHTRUDI, S., JOHNSON, M. K., & CHROSNIAK, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology & Aging*, **5**, 119-126.
- HENKEL, L. A., & FRANKLIN, N. (1998). Reality monitoring of physically similar and conceptually related objects. *Memory & Cognition*, **26**, 659-673.
- JOHNSON, M. K., FOLEY, M. A., & LEACH, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, **16**, 337-342.
- JOHNSON, M. K., FOLEY, M. A., SUENGAS, A. G., & RAYE, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, **117**, 371-376.
- JOHNSON, M. K., HASHTRUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- JOHNSON, M. K., & RAYE, C. L. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory and belief* (pp. 35-86). Cambridge, MA: Harvard University Press.
- KOUTSTAAL, W. K., & SCHACTER, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory & Language*, **37**, 555-583.
- LINDSAY, D. S. (1994). Memory source monitoring and eyewitness tes-

- timony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 27-55). New York: Cambridge University Press.
- LINDSAY, D. S., & JOHNSON, M. K. (in press). False memories and the source monitoring framework: Reply to Reyna and Lloyd (1997). *Learning & Individual Differences*.
- LINDSAY, D. S., JOHNSON, M. K., & KWON, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, **52**, 297-318.
- LOFTUS, E. F. (1981). Mental morphosis: Alterations in memory produced by the mental bonding of new information to old. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 417-434). Hillsdale, NJ: Erlbaum.
- LOFTUS, E. F. (1997, September). Creating false memories. *Scientific American*, **277**, 71-75.
- MARKHAM, R., & HYNES, L. (1993). The effect of vividness of imagery on reality monitoring. *Journal of Mental Imagery*, **17**, 159-170.
- MARSH, R. L., & HICKS, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1137-1151.
- MITCHELL, K. J., & JOHNSON, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). New York: Oxford University Press.
- MITCHELL, K. J., & ZARAGOZA, M. S. (1996). Repeated exposure to suggestion and false memory: The role of contextual variability. *Journal of Memory & Language*, **35**, 246-260.
- RABINOWITZ, J. C. (1989). Judgments of origin and generation effects: Comparisons between young and elderly adults. *Psychology & Aging*, **4**, 259-268.
- TOUSIGNANT, J. P., HALL, D., & LOFTUS, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition*, **14**, 329-338.
- WINOGRAD, E. (1968). List differentiation, recall, and category similarity. *Journal of Experimental Psychology*, **78**, 510-515.
- ZARAGOZA, M. S., & LANE, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 934-945.
- ZARAGOZA, M. S., & LANE, S. M. (1998). Processing resources and eyewitness suggestibility. *Journal of Legal & Criminological Psychology*, **3**, 305-320.
- ZARAGOZA, M. S., LANE, S. M., ACKIL, J. K., & CHAMBERS, K. L. (1997). Confusing real and suggested memories: Source monitoring and eyewitness suggestibility. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. Brainerd (Eds.), *Memory for everyday and emotional events* (pp. 401-425). Mahwah, NJ: Erlbaum.
- ZARAGOZA, M. S., & MITCHELL, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science*, **7**, 294-300.
- ZARAGOZA, M. S., MITCHELL, K. J., & DRIVDAHL, S. B. (1997). Imagery and false memory creation. Abstract reprinted in J. D. Read & D. S. Lindsay (Eds.), *Recollection of trauma: Scientific evidence and clinical practice* (p. 585). New York: Plenum.

NOTE

1. A previous experiment (Zaragoza & Mitchell, 1996, Experiment 1) demonstrated that placement of the single-exposure suggestions (i.e., whether in the first, second, or third subset of questions) had no effect on any of the dependent measures. Thus, to keep counterbalancing groups manageable, in the present study, all single-exposure suggestions were embedded in the last set of questions.

APPENDIX

Examples From the Postevent Questionnaires for All Conditions in Experiment 1A

High-Narrative-Overlap Condition

1. At the beginning of the scene, a young man dressed in jeans, a t-shirt, and *gloves* entered the house. Did he enter through the door?
2. The young man then walked into the bedroom, *pulled the window shade*, and went to the nightstand. Was the nightstand directly next to the bed?
- .
- .
- .
8. During the phone call the neighbor made to the police she said, "This is *Mrs. Anderson*, I'd like to report what I think is a robbery." She said she had gotten part of the license plate number. Did she say it was "HGN4073"?
9. Later was a scene in which two officers sitting in a police cruiser spotted the thief's car. When the officer set down his *coffee* and said, "Damn if it isn't!" did he suggest they follow the car?
- .
- .
- .
25. OK, returning once again to the beginning of the scene, a slender young man wearing jeans, a t-shirt, and *gloves* entered the house. Did he wear a jacket?
26. The thief later entered the bedroom, *pulled the windowshade down*, and began to rummage through the nightstand. Did he take anything from the nightstand?
- .
- .
- .
32. When the neighbor phoned the police, as she looked out the window she said, "This is *Mrs. Anderson*, I'd like to report what I think is a robbery." She gave details of the situation to the dispatcher. Did she appear confident?
33. In the next scene, there are two police officers sitting in a cruiser. One of the officers said he thought the car they'd seen was the one involved in the burglary. The other officer set down his *coffee* and responded, "Damn if it isn't!" Did he recognize the car by the license number?
- .
- .
- .

APPENDIX (Continued)

Moderate-Narrative-Overlap Condition

1. A young man wearing *gloves* entered the house. Did he enter through the door?
2. The young man *pulled the windowshade* and went to the nightstand. Was the nightstand by the bed?
.
.
.
8. The neighbor said, "This is *Mrs. Anderson*, I'd like to report . . ." Did she say the license number was "HGN4073"?
9. The officer recognized the car and set down his *coffee*. Did he suggest they follow it?
.
.
.
25. A young man wearing *gloves* entered the house. Did he wear a jacket?
26. The thief pulled the *windowshade* and looked through the nightstand. Did he take anything?
.
.
.
32. The neighbor said her name was *Mrs. Anderson* and that she wanted to report a robbery. Did she appear confident?
33. The officer spotted the car and set down his *coffee*. Did he recognize it by the license number?
.
.
.

Moderate-Narrative-Overlap + Irrelevant Information Condition

1. OK, if you are ready, here is the statement and question. A young man wearing *gloves* entered the house. Did he enter through the door?
2. If you are ready we will move on to the next one. The young man pulled the *windowshade* and went to the nightstand. Was the nightstand by the bed?
.
.
.
8. Please keep in mind that we are only asking you for a "yes" or "no" response as you read this question about the video. The neighbor said, "This is *Mrs. Anderson*, I'd like to report . . ." Did she say the license number was "HGN4073"?
9. Respond to the following question by answering "yes" or "no" only. We do not need to know exactly what was said. The officer recognized the car and put down his *coffee*. Did he suggest they follow it?
.
.
.
25. You should be able to answer each of these questions with a couple of words. A young man wearing *gloves* entered the house. Did he wear a jacket?
26. It is important that you read all the questions completely. The thief *pulled the windowshade* and looked through the nightstand. Did he take anything?
.
.
.
32. Please be sure to work at a steady pace but don't worry if others finish before you. We will continue when everyone has finished. The neighbor said her name was *Mrs. Anderson* and that she wanted to report a robbery. Did she appear confident?
33. I hope you remember that you are supposed to read each question carefully, try to remember the answer based on the video, and be sure to write your answer on the answer sheet provided. The officer spotted the car and set down his *coffee*. Did he recognize it by the license number?
.
.
.

Low-Narrative-Overlap Condition (the questions were presented in random order)

1. The thief checked his *gun* and went out. Did he slam the door?
2. The driver *jumped the curb* and said, "That's it!" Did the car stop?
.
.
.

APPENDIX (Continued)

8. When the police said, "Pull over or *we'll shoot!*" did the thief curse at the driver?
9. A young man wearing *gloves* entered the house. Did he enter through the door?
.
.
.
25. The thief pulled the *windowshade* and looked through the nightstand. Did he take anything?
26. The driver *smoked a cigarette* and waited. Did he appear anxious?
.
.
.
32. The thief checked his *gun* and walked out. Was there a porch?
33. The neighbor said her name was *Mrs. Anderson* and that she wanted to report a robbery. Did she appear confident?
.
.
.

The examples show only two iterations of each item; participants received three. Note that although no single participant would have received suggestions in all of these questions, they have been added here for the reader's information. Critical items were not italicized for participants; they are italicized here for the reader's convenience.

(Manuscript received August 3, 1999;
revision accepted for publication January 27, 2001.)