Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language xxx (2003) xxx–xxx
Journal of Memory and Language
www.elsevier.com/locate/jml

Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another event☆

D. Stephen Lindsay,[a,*] Bem P. Allen,[b] Jason C.K. Chan,[c] and Leora C. Dahl[a]

[a] *Department of Psychology, University of Victoria, Vic., Canada*
[b] *Department of Psychology, Western Illinois University, IL, USA*
[c] *Department of Psychology, Washington University, USA*

Received 30 May 2003; revision received 19 August 2003

Abstract

We explored the effect of the degree of conceptual similarity between a witnessed event and an extra-event narrative on eyewitness suggestibility. Experiments 1A and 1B replicated Allen and Lindsay's (1998) finding that subjects sometimes intrude details from a narrative description of one event into their reports of a different visual event. Those experiments also showed that intrusion rates were even higher when the narrative described the visual event itself. Experiment 2 replicated those findings, but found no more intrusions from a thematically similar versus dissimilar narrative. In Experiment 3 we disguised the relationship between the narrative and visual event, and obtained more intrusions from a thematically similar than dissimilar narrative. In Experiment 4 we obtained a thematic similarity effect when the relationship between narrative and visual event was disguised, but none when it was not. Results are discussed from the perspective of the source-monitoring framework.

Keywords: Memory; Eyewitness suggestibility; Misinformation effect; Source monitoring

Introduction

In Loftus's eyewitness misinformation paradigm, participants view a visual event, receive verbal information about that event that includes misleading suggestions, and are later tested on their memory of the event. Under a wide variety of conditions, participants sometimes err by reporting details suggested in the verbal information on tests of their memories of the witnessed event. For example, participants in a study by Loftus, Miller, and Burns (1978) viewed a slide show depicting an accident in which a car hit a pedestrian; for some participants, one slide showed the car at an intersection marked by a stop sign but a subsequent misleading question implied that the intersection had been marked by a yield sign. On a later memory test, such participants often falsely indicated that they had seen a yield sign in the slides.

In the 1980s, research on the misinformation effect emphasized the question of whether or not suggested details impair ability to remember the corresponding event details (memory impairment) (e.g., McCloskey & Zaragoza, 1985). From this perspective, the misinformation effect can be construed as an extension of earlier research on interference phenomena in verbal learning

☆ This research was supported by a research grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to the first author, and by a NSERC postgraduate scholarship to the fourth author. We thank Larry L. Jacoby and two anonymous reviewers for helpful critiques of an earlier version of the manuscript.

* Corresponding author. Fax: 1-250-721-9829.
E-mail address: slindsay@uvic.ca (D. Stephen Lindsay).

0749-596X/$ - see front matter © 2003 Published by Elsevier Inc.
doi:10.1016/j.jml.2003.08.007

studies (e.g., Keppel & Underwood, 1962; McGeoch, 1936; McGovern, 1964; Postman & Stark, 1969; for a more recent treatment of the memory-impairment issue in the context of the misinformation effect, see Chandler, Gargano, & Holt, 2001).

In the 1990s, the interest of researchers exploring the misinformation effect shifted to the question of whether participants believe that they are remembering something they witnessed when they report a suggested detail, as opposed to knowingly relying on the verbal information (either because they assume it is a reliable source of information about the event or because they think the experimenter wants them to report details from the verbal information). The issue of whether subjects sometimes confuse memories of suggestions as memories of the witnessed event can be cast as an extension of earlier verbal-learning research on list differentiation (e.g., Abra, 1972).

It is clear that participants in misinformation studies sometimes knowingly use memories of extra-event information when answering questions about a target event, but it is also clear that participants sometimes falsely believe they remember witnessing details that were merely suggested to them. For example, when conditions are such that it is easy for participants to differentiate between memories of the event and memories of the verbal information, informing participants immediately before the test that the verbal information was false and should not be reported eliminates the suggestibility effect, but when conditions make it difficult for participants to differentiate between memories from the two sources, such instructions attenuate but do not eliminate the suggestibility effect (Lindsay, 1990; Lindsay, Gonzales, & Eso, 1995; see also Holliday & Hayes, 2000, 2001, 2002; Roediger III, Jacoby, & McDermott, 1996).

In most versions of the misinformation paradigm, the misleading suggestions are embedded in verbal information *about* the witnessed event. For example, participants might read or hear a narrative summary of the event that includes misleading suggestions, or they might answer questions about the event that include inaccurate suppositions (e.g., "How fast was the car going when it passed the barn?" when there was no barn). Allen and Lindsay (1998), however, found that participants sometimes intruded details from a verbal description of one event into their memory reports of a different event. In that study, participants viewed a slide show of one event (e.g., a male professor conversing with a female student in a lecture hall) and read a narrative describing a different event (e.g., a female professor conversing with a male student in an office). The two events were unambiguously different, but each included certain common details (e.g., a can of pop was shown in the slides, and a different can of pop was mentioned in the narrative). Participants were tested on their memory of the visual event, with instructions that their reports should be restricted to things seen in that event. The key finding was that participants sometimes intruded details from the narrative into their reports of the slides. Presumably, these errors occurred because (a) the test questions were good cues to memories of the narrative (e.g., due to semantic overlap between questions and suggestions) and (b) participants sometimes accepted memories that popped to mind at test as accurate answers (e.g., because those memories were plausible answers).

One reason for interest in the misinformation effect is that it may generalize to real-world cases in which eyewitnesses are asked questions about a consequential event. To the extent such effects do generalize, the Allen and Lindsay (1998) findings raise the possibility that eyewitnesses' testimony may be compromised not only by suggestions regarding the witnessed event but also by memories of other events. Such information might come from a variety of sources (e.g., TV, books, other personal or vicarious experiences). It is also worth noting that deliberate uses of extra-event information are unlikely when that information concerns a different event (because, unlike in standard misinformation paradigms, there is no reason for subjects to assume that the narrative is a source of accurate information about the to-be-recalled event).

The current paper has two major aims. First, we sought to replicate the findings of the Allen and Lindsay (1998) experiment (which is the only extant demonstration of intrusions from a narrative description of one event into eyewitness reports of another event) with a variety of materials. Second, we assessed the role of the degree of conceptual similarity between the witnessed event and the narrative in modulating the rate of such intrusions. In the Allen and Lindsay experiment, the two sources were thematically similar (i.e., both concerned an interaction between a professor and a student). The SM framework implies that, all else being equal, intrusions should increase as a function of the conceptual similarity between the two sources (a prediction also made by earlier theories; e.g., Abra, 1972; McGeoch & McGeoch, 1937). Interestingly, however, prior studies on the effects of various kinds of source-similarity manipulations on the misinformation effect (reviewed in more detail below) have produced inconsistent results. Across the five experiments reported here, we explored conditions under which the degree of conceptual similarity between the target event and narrative does versus does not affect the rate of intrusions. Before reporting the experiments, we describe the theoretical context in which we pursued this work and summarize prior studies of the relationship between source similarity and suggestibility.

The source-monitoring framework

The source-monitoring (SM) framework holds that all experiences of remembering involve inferential pro-

cesses by which people attribute thoughts, images, and feelings to particular episodes in their personal past experience (for reviews, see Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 2000; for closely related views, see Jacoby, Kelley, & Dywan, 1989a; Whittlesea, 2003). SM attributions are thought to be based on the qualitative and quantitative characteristics of the content of accessed memory information, in the context of the testing situation. The notion is that memory information is attributed to particular sources in a way that is analogous to how current perceptual information is attributed to particular sources. For example, if a friend calls you on the phone, you may immediately recognize the person by his or her voice (perhaps also influenced by other sorts of knowledge, such as your expectancy that so-and-so is likely to call at a particular time of day). So too, if you recollect an utterance you may attribute it to a past experience of hearing someone speak, and more specifically you may attribute the statement to your friend Jean because the accessed memory information includes information about the sound of Jean's voice and/or because the semantic content of the utterance is in some way characteristic of Jean.

Source attributions can be made at a wide range of degrees of precision (e.g., from "I once read somewhere..." to "Yesterday afternoon at the checkout counter at Safeway I noticed the *National Enquirer*..."), depending partly on the rememberer's current goals. Like ongoing perceptual attributions, most memory-source attributions are made quickly and without conscious deliberation, but people sometimes have difficulty identifying source at the desired level of specificity (in which case they may consciously strive to retrieve additional source-specifying information and use analytic strategies to identify the source of memory information).

Source attributions are usually accurate, but sometimes they are faulty (e.g., you may think that Jean said something you had actually heard from Francis; that something occurred yesterday when it was really the day before; that you turned the oven off when you had merely thought about doing so; that you came up with an idea that you actually read, etc.). All else being equal, memory-source attributions are easier (and errors less likely) when the accessed memory information is uniquely characteristic of its true source. For example, if you have two friends with similar voices, you may be liable to mistake memories of an utterance by one as having been made by the other (just as you might mistake their voices on the phone). So too, as the amount of accessed memory information declines its source diagnosticity typically declines as well, such that SM tends to be more difficult if the event was not closely attended to during encoding, if there is a long delay between encoding and test, or if memory test conditions interfere with retrieval (akin to not recognizing a voice on the phone when the connection is bad).

Source-monitoring accuracy is also affected by the appropriateness and stringency of the decision-making processes and criteria used. Under many conditions, people largely disregard the sources of the thoughts and images that come to mind (because source discrimination is not relevant to the task at hand; e.g., if the goal is to solve a problem one may not attend to the memorial sources of the potential solutions that come to mind). Under other conditions, people make source attributions quickly and automatically, and under yet others they set stringent criteria and scrutinize memory information systematically before making a source attribution. All else being equal, SM accuracy increases with the stringency and appropriateness of the criteria and processes used (although errors may sometimes occur even with stringent and appropriate criteria, e.g., if the two sources are highly similar).

As in many areas of cognitive psychology, the various factors hypothesized to influence SM have usually been investigated *cetaris paribus* (Palmer, 1999, cited in Vecera, Vogel, & Woodman, 2002). Thus at this stage in its development the SM framework makes few predictions regarding interactions between the parameters thought to affect source monitoring (e.g., dimensions of similarity [visual, auditory, conceptual, etc.], expectations, aspects of source [where, who, how, when?], the precision with which sources are identified, and the kinds and stringency of source-monitoring criteria). Manipulating a particular dimension of source similarity, for example, may have no effect if, at test, participants make their source attributions on other dimensions or do not attend to source at all (see Marsh & Hicks, 1998).

Eyewitness suggestibility and the SM framework

The SM framework has been developed and explored using a variety of research procedures, including variants of the misinformation paradigm. Here we briefly summarize prior studies of eyewitness suggestibility that were inspired by (and informed further developments of) the SM framework.

Lindsay and Johnson (1989b) demonstrated a standard suggestibility effect in a procedure in which exposure to verbal suggestions preceded (rather than followed) witnessing the visual event (see also Abeles & Morton, 1999; Holliday & Hayes, 2002). Those findings are consistent with the idea that misinformation effects can reflect confusions between different memory sources (as opposed to reflecting a mechanism by which new information updates memory representations of prior information).

Lindsay and Johnson (1989a) found that a suggestibility effect obtained with a yes/no recognition test was absent among participants tested on a SM test that required them to specify the source (visual event, narra-

tive, both, or neither) of each test item (see also Zaragoza & Koshmider, 1989). Lindsay and Johnson interpreted these findings as evidence that the SM test encouraged participants to attend to source-specifying memory information, thereby enhancing their ability to differentiate between memories of the target visual event versus memories of the verbal information. Converging evidence for that conclusion came from Zaragoza and Lane (1994, Experiment 2), in which suggestibility effects were reduced but not eliminated by a SM test as opposed to yes/no test (see also Chambers & Zaragoza, 2002; Hekkanen & McEvoy, 2002; for an interesting exception, see Hicks & Marsh, 2001).

Zaragoza and Lane (1994) found that presenting suggestions in ways that encourage participants to encode them elaboratively and in integration with event memories increased reports of suggestions on a subsequent SM test (see also Mitchell & Zaragoza, 1996). Lindsay (1990) adapted Jacoby's (e.g., Jacoby, Woloshyn, & Kelley, 1989b) opposition procedure, by correctly informing participants before a cued recall test that any potential answer that had been mentioned in the narrative was a misleading suggestions and hence should not be reported: Consistent with the SM framework, under conditions that made it easy for participants to differentiate memories of the narrative from memories of the event, these instructions enabled participants to avoid reporting suggestions, but under conditions that made SM difficult participants often intruded suggestions despite the opposition instructions (see also Holliday & Hayes, 2000, 2001, 2002; Lindsay et al., 1995; Roediger et al., 1996). Eberman and McKelvie (2002) found higher rates of reports of suggestions on a SM test among participants with high imagery than among those with low imagery. Frost, Ingraham, and Wilson (2002) found that, as predicted by the SM framework, source accuracy declined more steeply than recognition memory over a 1-week delay.

A key premise of the SM framework is that, all else being equal, increasing the similarity between memories of suggestions and memories of the witnessed event will increase false reports of suggestions. Several studies cited above support that conclusion, but there are also published reports of null effects of source-similarity manipulations on suggestibility.[1] Bonto and Payne (1991) found no significant effect of same versus different environmental context during the event and postevent information on performance on the standard Loftus 2-alternative forced-choice (2AFC) test (although the means were in the predicted direction, with a 37% difference in reports of suggestions on misled vs. control items in the same-context condition as opposed to only a 22% difference in the different-context condition). In two studies with bilingual participants, Shaw III, Garcia, and Robles (1997) reported null effects of same versus different language during the event and misinformation phases on performance on a subsequent test (2AFC or cued recall). More recently, Mitchell and Zaragoza (2001) reported two experiments in which they manipulated the extent to which the misleading narrative recapitulated the event. Suggestibility effects on a SM test were reliable and of the same magnitude regardless of whether the narrative was a detailed and coherent summary of the event or a relatively disjointed and brief series of statements ordered differently from the sequence in the event itself.

Why do source-similarity manipulations modulate the size of the misinformation effect under some conditions but not others? One possibility is that the inconsistency has to do with variations in the extent to which deliberate uses of extra-event information contribute to apparent suggestibility effects. As noted above, participants sometimes select or report items that they know they recollect encountering in the extra-event material, either because they assume that material is reliable or because they think the experimenter wanted them to rely on that material. Aware uses of extra-event information are especially likely when that information is apprehended as being about the to-be remembered event and is presented shortly before the test and/or processed in ways that support good memory for the suggestions, and when the test instructions do not warn participants against relying on the extra-event information. To the extent that reports of suggested items reflect aware or deliberate uses of extra-event information, there is no reason to expect that source-similarity manipulations would affect the rate of such errors.

A related point is that standard 2AFC tests (like that used by Bonto & Payne, 1991) may encourage participants to use a familiarity criterion for response selection. Because most test items pair a witnessed detail with a novel foil, participants may slip into a tendency to select whichever response "pops out" as familiar (unless they are explicitly warned against doing so) (see Lindsay & Johnson, 1989b). Here again, to the extent that participants base their responses on familiarity without regard to source there is little reason to believe that source-similarity manipulations would be effective.

It is also likely that different sorts of source-similarity manipulations have differential effectiveness. For example, participants in the Shaw et al. (1997) studies were fluent Spanish–English bilinguals living in El Paso, Texas; such individuals are likely to have considerable practice at abstracting meaning away from its linguistic medium (and in any case the witnessed event was pre-

[1] Null effects of source-similarity manipulations have also been reported in several experiments using McCloskey and Zaragoza's (1985) modified test, but we do not discuss such findings here because the modified test precludes subjects from reporting suggestions and hence is likely to be insensitive to source-similarity manipulations (Lindsay & Johnson, 2001).

sented silently; only the instructions that preceded and followed it were in one language or another), so the similarity of memories of the event versus the extra-event information may not have differed much as a function of same versus different language. Likewise, the manipulation of the extent to which the narrative recapitulated the event in the Mitchell and Zaragoza (2001) studies may have had little effect on the extent to which memories of particular details from the narrative were similar to memories of details from the event: In both cases, participants likely apprehended the narrative as being about the event. In the absence of a validated model of the psychological construct of similarity, these are merely speculations, but in our judgment they are plausible ones.

With these considerations in mind, in the experiments below we used strong similarity manipulations, test instructions that clearly specified that responses were to be based solely on memory for the visual event, and cued recall tests rather than 2AFC.

Experiments 1A and 1B

Experiments 1A and 1B had three major aims. First, we sought to replicate, with a new set of materials, Allen and Lindsay's (1998) finding that participants sometimes intrude details from a narrative describing one event into their reports of another witnessed event. Second, we went beyond the Allen and Lindsay (1998) procedure by decreasing the thematic similarity between the two events: Whereas in Allen and Lindsay's study the witnessed event and narrative both involved student-professor conversations, in the present experiments the witnessed event involved searching a house for a misplaced object and the narrative involved housecleaning. Third, we compared the rates of intrusions from a different-event narrative versus from a narrative describing the visual event itself.

From the perspective of the SM framework, memories of a narrative description of a witnessed event should be more similar to memories of that event itself than are memories of a narrative description of a different event. This is because semantic content is an important cue to source (e.g., Lindsay, Johnson, & Kwon, 1991, Experiment 3). Information that is apprehended and encoded as being about the witnessed event should, from this perspective, give rise to memories that are both (a) likely to come to mind during attempts to recollect relevant aspects of the to-be-remembered event and (b) likely to be accepted as memories of the event itself. Thus we predicted that the misinformation effect would be larger in the same-event narrative condition than in the different-event narrative condition.

When the misleading narrative describes the target event, there is a possibility that participants will knowingly use memories of the narrative as a source of answers at test. We sought to reduce that possibility by instructing participants to base their answers solely on what they remembered seeing in the target event. To provide converging evidence as to the extent to which participants knowingly relied on the narrative as a source of answers, Experiments 1A and 1B also included conditions in which the test instructions warned participants that the narrative included details not present in the slides. This is a relatively weak warning (much weaker than the opposition instructions, which tell participants that any relevant detail remembered from the narrative is wrong and should not be used as a basis for responding), but we reasoned that if participants in the non-warned conditions were knowingly using memories of the narrative as a basis for answers then this warning would reduce such aware uses of the extra-event information.

The experiments also compared a Web-based version of the procedure (Experiment 1A) with a laboratory procedure (Experiment 1B). Prior studies indicate that Web research can yield results comparable to "live" research (Allen, 1999; Birnbaum, 2000; McGraw, Tew, & Williams, 2000), but to our knowledge no previous study has compared Web-based and laboratory versions of a misinformation procedure.

Method

Participants

There were 303 participants in Experiment 1A; as explained below, 15 of these were dropped due to poor performance on easy filler items on the memory test, so analyses are based on the remaining 288 (57% women). Most (56%) were college/university students who earned credit toward a course at a college or university for their participation, but others were people of a variety of ages who spontaneously volunteered to participate (age range from 14 to 56 years, $M = 24.81, SD = 9.80$). Most participants (85%) were from the United States, but countries around the world were represented. Participants logged on to the experiment through various research sites (e.g., the American Psychological Society site).

The participants in Experiment 1B were 146 undergraduate students at Western Illinois University. Two participants were dropped due to poor performance on easy filler questions on the memory test, so analyses were based on the remaining 144 participants (64% female, age range from 17 to 23 years, $M = 19.03$, $SD = 1.30$). They received credit points in an introductory psychology class in exchange for participating.

Materials and procedure

The materials and procedures in the two experiments were essentially the same, but Experiment 1A was conducted on the web whereas Experiment 1B was

conducted in a laboratory with groups of 2–11 participants and a "live" experimenter. For Experiment 1B, the web pages used in Experiment 1A were printed (with the slides in color, as on the web site) and distributed to participants in booklet form (with pages in the same order as that of the web pages), and participants wrote in the buttons and dialog boxes with a pencil. Because the procedures were otherwise identical, below we provide details only of Experiment 1A.

The first web page provided a brief description of the study and informed consent information, followed by a text describing the target event illustrated with color slides. The target event involved a man searching his house to find a misplaced gold coin. As his search progressed through different rooms, he encountered six clusters of from 9 to 14 objects: tools (including a saw and plumber's wrench); beauty products (tweezers and mirror); small ceramic animals (mouse and rabbit); a place setting (plate and mug); home office supplies (magic marker and stapler); and home hardware (bolt and nail). Each cluster was depicted in a separate slide, and the objects in parentheses were mentioned in the accompanying narrative (e.g., "On the dressing table top he found an assortment of hair-care and make-up items. There were some tweezers in the middle of the items arrayed on the table top. On the far left edge of these items there was a large hand mirror"). We refer to the objects mentioned in the text (e.g., "mirror") as "marker objects," because objects not actually present in the slides—suggested lures—were "placed" next to them in a subsequent misleading narrative (e.g., "He picked up a cotton swab that was next to a mirror, then tossed it back onto the counter-top"). The lures for each cluster were as follows (List-A lure first): pliers, screwdriver (tools); cotton ball, cotton swab (beauty products); squirrel, raccoon (small ceramic animals); salt shaker, butter dish (place setting); eraser, pencil (home office supplies); drill bit, nut (hardware).

After reading the event narrative and looking at the slides, participants clicked a link to a page that contained three stories: beginning and ending filler stories and a second-position misleading narrative. Participants were assigned to one of two misleading narratives by alternating the link to them every few days. One narrative was *the same as* the event, in that it described the same "gold coin" scenario as in the initial event, with the same character searching the same house and encountering the six clusters of objects (with one marker object and one suggested object per cluster). The other narrative was *different from the event*: Although there were still certain parallels between the target event and the narrative, in that both involved a protagonist moving through a house and encountering various clusters of objects, in the different-event condition the protagonist was a teenage girl rather than a man, her goal was to complete house-cleaning chores, the house was not described as being the same as that in the target event, and the order in which the object clusters were encountered differed from that in the target event.

Participants were admonished on the narrative page that if they returned to the slide show their responses would be invalidated and their time, as well as that of the experimenter, would be wasted (a similar warning regarding returning to the stories was printed at the beginning of the next page, described below). After the participants read the stories, they clicked a link to a questionnaire page in CGI format that asked for consent to use their data and asked them to report age, sex, and place of residence.

Next, a series of filler questions focused on the filler stories to support the perception that all of the stories were important. Finally, a series of questions about the target-event slide show was preceded by a plea that participants answer the questions solely on the basis of their memories of the slides presented in the first phase of the experiment. They were also asked not to leave any question unanswered (although 38.5% of participants in Experiment 1A and 26.0% of those in Experiment 1B left unanswered one or more questions for which they had received suggestions). For approximately half of the participants, the test instructions warned that some of the details mentioned in the second story were NOT in the slides, and "you cannot assume that just because something was mentioned in the story it is a correct answer to the test questions about the slide show." The remaining participants did not receive this warning. Participants were assigned to the "warning" and "no warning" conditions by changing the link from the narrative page to the final questionnaire every few days.

There were two critical cued-recall questions regarding each object cluster (one for each "marker" object), with one question asking about a control item and the other asking about a suggested item (e.g., "What was near the outer, rounded edge of the hand mirror that was on the dressing table top?" was a control question for participants assigned to Lure Set A [who had read no suggestion regarding something close to the mirror] and an experimental question for participants assigned to Lure Set B [who had read a suggestion that a cotton swab was close to the mirror]). Intermingled among the critical questions were four "easy" questions regarding details in the slide show that anyone who had viewed the slides would likely remember. These questions were included to establish that participants had the slide show in mind when answering questions about it. As mentioned earlier, participants who made more than one error on the four easy questions (15 participants in Experiment 1A [5%] and 2 in Experiment 1B [1%]) were excluded from all analyses.

The slides can be seen at http://www.wiu.edu/users/mfbpa/slideshow.html. The stories can be viewed at the same URL, substituting the following for "slideshow:"

gala2, galb2, guia1, guib1. The test is at http://www.wiu.edu/users/mfbpa/memoryquesab.html (substitute "memoryquesabw" for the warning condition).

Results

"Easy" questions

For the "easy" questions, mean proportion accurate was near ceiling (Ms = .974 and .897 in Experiments 1A and 1B, respectively), with no reliable difference across counterbalancing and warning/no-warning conditions (all Fs < 1).

Probability of false recall

In each experiment, the proportion of the six critical items on which participants reported suggested details was analyzed in a 2 (suggested vs. control) × 2 (same-event vs. different-event narrative) × 2 (warning vs. no warning) × 2 (counterbalancing A vs. B) analysis of variance (ANOVA), in which the first variable was a repeated measure. Effect size was indexed with partial eta squared (*es*; .01 or less is a small effect, .06 a moderate effect, and .14 and above a large effect). Descriptive data are depicted in Fig. 1.

In each experiment, there was a reliable suggestibility effect, as revealed by a significant effect of item type (suggested vs. control): in Experiment 1A, $F(1,280) = 97.44$, $MSe = .01305$, $p < .001$, $es = .258$, and in Experiment 1B, $F(1,136) = 48.12$, $MSe = .01747$, $p < .001$, $es = .261$. There were also reliable effects of the between-subjects manipulation of narrative type: in Experiment 1A, $F(1,280) = 19.00$, $MSe = .01242$, $p < .001$, $es = .064$, and in Experiment 1B, $F(1,136) = 18.71$, $MSe = .01805$, $p < .001$, $es = .121$. Most importantly, there was a reliable Item Type × Narrative Type interaction in each experiment: in Experiment 1A, $F(1,280) = 16.04$, $MSe = .01305$, $p < .001$, $es = .054$, and in Experiment 1B, $F(1,136) = 10.19$, $MSe = .01747$, $p < .001$, $es = .07$. No other main effects or interactions were significant (all Fs < 3.5, all ps > .069). Follow-up tests showed reliable suggestibility effects in both the same-event condition (Experiment 1A: $F(1,143) = 68.914$, $MSe = .01823$, $p < .001$, $es = .325$; Experiment 1B: $F(1,71) = 33.22$, $MSe = .02697$, $p < .001$, $es = .327$) and the different-event condition (Experiment 1A: $F(1,143) = 29.39$, $MSe = .007636$, $p < .001$, $es = .171$; Experiment 1B: $F(1,71) = 15.43$, $MSe = .007941$, $p < .001$, $es = .178$).

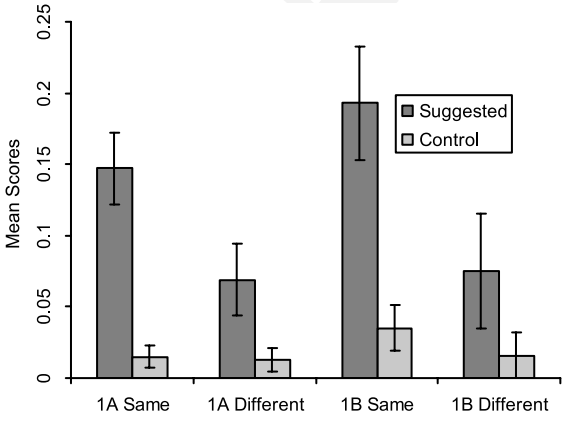

Fig. 1. Mean proportion false recall on suggested versus control items in Experiments 1A and 1B as a function of whether the narrative in which suggestions were imbedded described the same event as the witnessed event or a different event. Error bars are 95% confidence intervals calculated using the MSE from the Item Type × Narrative Type ANOVA (see Loftus & Masson, 1994).

Discussion

Results of the internet and laboratory versions of the procedure were nearly identical, supporting the viability of web-based experiments on the misinformation effect. Moreover, the Allen and Lindsay (1998) finding was replicated and extended with these new materials and procedures, in that participants in the different-event condition sometimes intruded details from their memories of the house-cleaning narrative into their reports of the gold-coin-search slides. The finding that a thematically distinct story can give rise to false reports dramatically extends the range of potential sources of contamination in eyewitness reports, even relative to the Allen and Lindsay study.

As predicted, the rate of false reports was substantially and significantly higher when suggestions were embedded in a narrative description of the witnessed event than when the same suggestions were presented in a narrative description of a different event. This finding is consistent with the SM framework: Memories of the narrative would be expected to be more similar to memories of the visual event in the same-event narrative condition than in the different-event condition, and hence participants in the same-event condition would be more likely to retrieve memories from the narrative at test and more likely to accept them as memories of the target event.

It is possible that participants in the same-event condition sometimes based responses on memories that they knew came only from the narrative (despite the test instructions to base reports only on the slides). Because there would be little reason for those in the different-event condition to base reports on memories they knew came only from the story (given that the story was clearly about a different event), such aware uses of story-only memories in the same-event condition may have contributed to the difference in intrusion rates in the two conditions. As evidence against that possibility, the null

effect of the warning suggests that participants in the same-event condition did not knowingly rely on the narrative as a source of answers (i.e., if they had, those warned that the narrative included inaccuracies should have less often relied on it).

We propose that memories of suggestions are more likely to come to mind and be accepted as memories of the event if the narrative is conceptually similar to the to-be-remembered event. The extreme of conceptual similarity is when the narrative is apprehended as a faithful account of the event itself, but we speculate that less extreme degrees of conceptual similarity can also affect the likelihood of source errors. If so, then such a conceptual source-similarity effect may also be observed when two different-event narratives are compared: One that is thematically similar to the visual event and one that is thematically dissimilar. Even though the surface form of the specific suggestions would be identical in these two sorts of narratives, the encoding of those suggestions should produce memories that are more likely to come to mind and be accepted at test if the suggestions were encountered in the context of a similar-event narrative than if they were encountered in the context of a dissimilar-event narrative. Because both narratives in this comparison describe a different event from the witnessed event, there would be equally little reason for participants to knowingly rely on narrative-only memories in each condition. Thus higher rates of intrusions from a similar-event narrative than from a dissimilar-event narrative would provide additional evidence for the SM framework. This issue was explored in Experiments 2–4.

Experiment 2

Experiment 2 was designed with two objectives in mind. First, we aimed to replicate the central findings of Experiments 1A and 1B (i.e., a reliable suggestibility effect both when the narrative described the witnessed event and when it described another event, but with that effect larger in the former condition) with a new set of materials. The event in Experiment 2 was a museum burglary, and it was presented as a video (rather than a series of pictures) and the narrative was presented auditorially (rather than as text). Second, we tested the hypothesis that intrusion rates would be higher when the event and narrative were thematically similar than when they were thematically dissimilar. Specifically, we compared performance in three conditions: One in which the narrative summarized the witnessed event itself (a burglary in a museum), another in which the narrative described a thematically similar but clearly different event (a burglary in a palace), and a third in which the narrative described a thematically dissimilar event (a school field trip to a palace).

Method

Participants

Participants were 185 undergraduate students (63% women) at the University of Victoria who volunteered for optional extra credit in an introductory psychology course. Participants were randomly assigned to six cells: same-event A ($n = 30$), same-event B ($n = 31$), similar-event A ($n = 32$), similar-event B ($n = 32$), dissimilar-event A ($n = 30$), and dissimilar-event B ($n = 30$). Age data were not collected.

Materials

An 8-min portion of the movie *Return of the Pink Panther* was used as the witnessed (video) event. The video contained a scene in which a man went through several different settings in a museum to perform a burglary. The original sound of the video was replaced by an accompanying audio narrative describing the events in the video, which was recorded by the same woman's voice as in the postevent narrative. The same-event narrative described the museum burglary shown in the witnessed event video. The similar-event narrative described a burglary that happened in a palace, whereas the dissimilar-event narrative described a school field trip to a palace. There were two versions of each of these narratives, which differed only with regard to which of two sets of four suggestions they included (Version A: "no smoking" sign, a mounted deer head, sculpted cupids, a fire escape ladder; Version B: a miniature building, a sculpted knight, fresco of Greek goddesses, cans of paint).The suggested details were selected to be plausible and imagable yet unlikely to be generated as guesses. As in Experiments 1A and 1B, the suggestions were of new details not actually witnessed in the event (rather than subtly different descriptions of witnessed objects).

An imagery-rating questionnaire asked participants to assess the vividness of their imagery while listening to the narrative. The imagery-rating questionnaire was part of a cover task used to encourage participants to imagine the people, objects, and actions described in the narrative, and its results will not be analyzed here.

The memory test questionnaire contained 12 cued-recall questions: 4 questions concerned details not present in the video but mentioned in the narrative (suggested items), 4 concerned details neither present in the video nor mentioned in the narrative (control items), and 4 concerned details obviously present in the video (easy items). In addition to a general instruction to "Answer the questions based only on what you saw in the video," each question reminded participants that they were being asked about the video, not the narrative. An example critical question is: "What was sculpted onto the stand in the big room in the video?" (in fact, no such sculpting appeared in the stand in the video, but the

narrative heard by some participants mentioned that there were cupids sculpted into the base of the stand).

A 10-min unrelated distracter task was presented during the delay period between the post-event narrative and the memory test. In this task, participants were asked to find as many differences as they could between two similar pictures.

Procedure

Participants were tested individually or in groups of two to four. After giving informed consent, participants viewed the museum burglary video in preparation for a memory test. The video was presented on a television with the accompanying narrative presented via a cassette tape. Participants listened to the postevent narrative immediately after viewing the video, and were asked to visualize the people, actions, and objects described in the narrative as vividly as possible. Next, participants rated their images and then completed the 10-min find-the-difference distracter task.

After the distracter task, participants were given the memory test. They were asked to base their responses solely on the video, and to answer every test question (although 31% left unanswered one or more questions for which they had received suggestions). Following the memory test, all participants were fully debriefed.

Results

Accuracy on filler questions

A between-subjects ANOVA revealed a reliable effect of narrative type on performance on the four easy questions, $F(2,182) = 8.065$, $MSe = .021$, $p < .001$, $es = .081$. Post hoc tests with Bonferroni adjustment showed that participants in the same-event condition ($M = .98$) outperformed participants in the similar-event condition ($M = .88$), $p < .01$, and the dissimilar-event condition ($M = .90$), $p < .01$, with the latter two conditions not differing, $p > .05$. The answers to the easy questions were not mentioned in the narrative, but it may be that spontaneous recall of those details while listening to the narrative was more common with the same-event narrative than with the other narratives.

Probability of false recall

Counterbalancing version (A vs. B) had no reliable main effect in any analysis, nor did it enter into any reliable interactions (all Fs < 1.65). Therefore, all subsequent analyses were collapsed across this variable.

The mean probabilities of false recall are depicted in Fig. 2. These data were analyzed in a ANOVA with item type (misinformation reported for the critical and control questions) as a within-subjects variable and narrative type (same, similar, and dissimilar) as a between-subjects variable. There were significant main effects

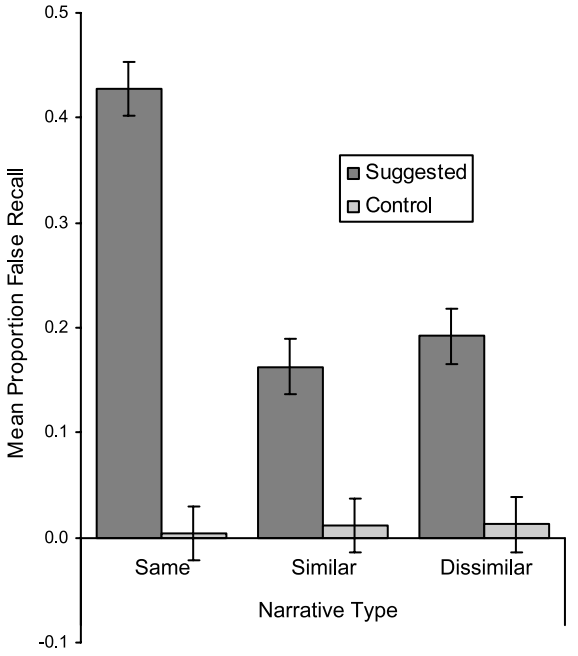

Fig. 2. Mean proportion false recall on suggested versus control items in Experiment 2 as a function of whether the narrative in which suggestions were imbedded described the witnessed event, a different but thematically similar event, or a different and dissimilar event. Error bars are 95% confidence intervals calculated using the MSE from the Item Type × Narrative Type ANOVA (Loftus & Masson, 1994).

of item type, $F(1,182) = 180.5$, $MSe = .032$, $p < .01$, $es = .498$, and narrative type, $F(1,182) = 20.24$, $MSe = .030$, $p < .01$, $es = .182$. The interaction between item type and narrative type was also significant, $F(2,182) = 21.55$, $p < .01$, $es = .191$. Because participants almost never spontaneously reported suggested details as guesses on control items (overall $M = .0095$), the subsequent simple effects tests were conducted only for the items for which participants had received suggestions. The probability of false recall in the same-event condition exceeded that in the similar-event condition, ($t = 6.01$, $SE = .044$, $p < .01$) and that in the dissimilar-event condition ($t = 4.99$, $SE = .047$, $p < .01$). However, contrary to our hypothesis, no reliable difference was found in the rate of false recall between the similar- and dissimilar-event conditions, $t = .71$.

Discussion

Results in the same-event and similar-event narrative conditions mirrored those of Experiments 1A and 1B: a reliable misinformation effect was obtained in both conditions, and that effect was substantially larger in the same-event narrative condition than in the similar-event

narrative condition. Contrary to expectations, however, the intrusion rate in the similar-event narrative condition was no higher than that in the dissimilar-event narrative condition. On the face of it, this null effect joins those reported by Bonto and Payne (1991), Shaw et al. (1997), and Mitchell and Zaragoza (2001) as evidence against the SM framework's prediction that, all else being equal, false reports should increase with source similarity. We conducted several versions of Experiment 2 (varying the materials and instructions across experiments), and consistently failed to find a reliable difference between similar- and dissimilar-event narrative conditions.

We had hypothesized that memories of suggestions encountered in a similar-event narrative should be processed in ways that make them more similar to memories of the event itself, and hence more likely to come to mind and be accepted as answers at test. Why was this not the case? Perhaps because these effects were attenuated by situational factors that led participants to link the event and narrative even in the dissimilar-narrative condition. In Experiment 2, the event, narrative, and test were all presented close together in time and as parts of a single experiment, conducted in the same place, and presented by the same experimenter. Also, despite having a different theme, the dissimilar-events narrative had considerable overlap with the visual event (e.g., both could be described as a tour through a fancy building filled with numerous art objects) and with the test questions. The concomitance between event and narrative, coupled with the overlapping semantic content between event and narrative and between test and narrative, likely made the relationship between the two sources patently obvious even to participants in the dissimilar events condition. In turn, the obviousness of the relationship between the two sources may have overwhelmed any effect of our manipulation of the thematic similarity between the narrative and the event on intrusion rates. That is, when event, narrative, and test are concomitant, participants are likely to apprehend the two sources as related even in the dissimilar-events condition, and hence there will be no effect of thematic similarity. Spencer and Weisberg (1986) reported a related effect on analogical transfer: When there was a delay between the training and target problems, transfer was much greater when the two phases were presented as parts of a single experiment than when the target problem was presented as an ostensibly unrelated lecture demonstration; when there was no delay, in contrast, transfer was equivalently high regardless of the context manipulation. If concomitance likewise trumps thematic similarity in our suggestibility procedure, then separating the narrative and the event so as to disguise the relationship between the two should yield an effect of thematic similarity on intrusion rates.

Experiment 3

The purpose of Experiment 3 was to compare intrusion rates from similar- and dissimilar-event narratives under conditions that disguised the relationship between the witnessed event and the narrative. To that end, the narrative and video were presented to participants in the context of two different experiments, separated by 24 h and conducted in different rooms and by different experimenters. We presented the narrative in Session 1 and the video and memory test in Session 2 (cf. Abeles & Morton, 1999; Holliday & Hayes, 2002; Lindsay & Johnson, 1989b, for prior examples of such "reversed" misinformation effect procedures). We used this temporal order for three reasons: First, presenting the narrative immediately before the test (as in Experiments 1 and 2) likely reduces effects of thematic similarity (because regardless of similarity the recency of the suggestions would make them highly accessible); a substantial delay between the narrative and test creates opportunities for similarity (rather than recency) to influence remembering. Second, the delay between narrative and test would likely reduce the likelihood that subjects spontaneously identify the source of memories of the narrative when they come to mind (simply because source monitoring declines with delay). Third, the reversed procedure eliminates the possibility that, while listening to the narrative, participants would spontaneously note differences between the narrative and the witnessed event (which might serve to reduce misinformation effects; cf. Tousignant, Hall, & Loftus, 1986).

Method

Design

In Session 1, participants listened to an audio narrative. In Session 2 (conducted 24 h later, in a different room and with a different experimenter) participants viewed the video and took the memory test. The narrative presented in Session 1 was either thematically similar or thematically dissimilar to the witnessed event video. Two versions of each type of narrative were created, differing in terms of which of two sets of six suggestions they contained. Thus there were four cells in the design: similar-event A ($n = 16$), similar-event B ($n = 16$), dissimilar-event A ($n = 16$), dissimilar-event B ($n = 19$).

Unlike Experiments 1A, 1B, and 2, there were no true control questions on the memory test. Rather, each participant was exposed to one of two suggestions regarding each of six critical items. The guessing baseline was the number of questions on which each participant guessed suggested details from the non-presented narrative divided by the number of questions on which that participant did not report suggested details from the presented narrative. This method enabled us to increase the number of suggestions given to each participant;

note that Experiment 2 established that the guessing baseline for these items was near zero.

Participants

Participants were 67 undergraduate students (69% women) at the University of Victoria, who volunteered for optional extra credit in an introductory psychology course or for $10 ($5 for each session). Age data were not collected. Participants were randomly assigned to the four cells of the design and were tested individually or in groups of two to three.

Materials

As in Experiment 2, the similar-event narrative described a palace burglary, whereas the dissimilar-event narrative described a school field-trip to a palace. Each narrative was approximately 1200 words long, and included six suggestions. Version A and Version B of the narratives differed only in which of the six suggestions they contained, and these suggestions were counterbalanced across participants (Version A: "no smoking" sign, painting of a prince, statue of a Greek goddess, medieval swords, sculpted dolphins, fresco of knights in armor; Version B: "no photography" sign, painting of a queen, statue of a man wearing armor, medieval shield, sculpted cupids, fresco of David and Goliath). The suggestions were of new objects, not present in the video. The witnessed event video was the same video clip of a museum burglary used in Experiment 2. In the current experiment, however, the video clip and accompanying narrative were digitized. An additional filler video was used as part of the cover task for Session 2. This filler video was a 9-min excerpt (with the original sound track) from the IMAX movie *Everest*, which showed hikers climbing Mount Everest. None of the critical details in the witnessed event or the narratives appeared in the filler video. Both videos were presented on a 15-in. computer screen with external speakers.

Three questionnaires were used for this experiment: An auditory perception questionnaire, a visual perception questionnaire, and a memory test questionnaire. The first two questionnaires were part of the cover tasks and results from them will not be reported. The auditory perception questionnaire asked participants to rate their perceptions of the auditory components of the two videos. The visual imagery questionnaire asked participants to rate their imagery while listening to the narrative. The memory test questionnaire consisted of nine cued recall questions like those used in Experiment 2. Six questions concerned the critical details presented only in the narrative, while the other three questions were easy filler questions regarding details obvious in the video.

Procedure

The experiment consisted of two sessions scheduled 24 h apart. Session 1 was conducted by a woman in a room on the third-floor of the psychology building. After giving informed consent, participants listened to the misinformation narrative, with instructions to visualize what was described in the narrative. Afterwards, participants rated how vivid their images were. Participants then listened to the narrative again, with instructions to visualize it in vivid colors, and then rated how colorful their images were. The experimenter then provided a bogus debriefing, according to which the experiment was designed to determine whether people imagine events in color spontaneously and whether they can imagine in color when instructed to do so. The true purpose of the imagery instructions and the repeated presentation of the narrative was to enhance participants' memory for the details in the narrative. All participants were awarded credit or $5 at the conclusion of Session 1.

Session 2 was conducted by a man in a room in the basement of the psychology building. Participants were told that they would view a video clip depicting a burglary, with the original soundtrack removed and replaced by a woman's voice narrating the action, followed by a second video clip about mountaineering with its original soundtrack. After participants viewed the critical video, they rated the loudness of the event depicted in the video. Participants then viewed the filler video and answered the same question about that video (as part of the cover task). Finally, participants were given the surprise memory test. Participants were told to base their responses solely on what they saw on the video, and to answer every question (although 28.3% left one or more questions unanswered). Following the completion of the memory test, participants were thoroughly debriefed and awarded their credit or $5 for Session 2.

Results

Accuracy on filler questions

Performance on the filler questions was near ceiling and did not reliably differ between the similar-event condition ($M = .96$) and the dissimilar-event condition ($M = .92$), $t = 1.1$.

Probability of false recall

In the current experiment, participants who listened to narrative version A served as the control group for participants who listened to narrative version B, and vice versa. Therefore, if a participant who heard narrative version A spontaneously reported a critical detail that was mentioned only in narrative version B, it was counted as false recall in the control condition (and indexed as the proportion of questions on which suggested details were not reported that such guesses were reported). Counterbalancing version (A vs. B) had no reliable main effect in any analysis nor did it enter into any

reliable interactions (all Fs < 1.61). Therefore, all subsequent analyses were collapsed across this variable.

The mean rates of false recall are depicted in Fig. 3. These data were analyzed in a ANOVA with narrative type (similar vs. dissimilar) as a between-subjects variable and item type (critical vs. control) as a within-subjects variable. Participants more often reported a critical misleading detail when they had heard it in the narrative than when they had not, $F(1,65) = 80.94$, $MSe = .021$, $p < .01$, $es = .56$. Those in the similar-event condition recalled more misinformation than did those in the dissimilar-event condition, $F(1,65) = 7.83$, $MSe = .026$, $p < .01$, $es = .108$. The interaction between narrative type and item type was also significant, $F(1,65) = 8.58$, $p < .01$, $es = .12$. Fisher's protected simple effects t-tests of the rates of false recall of suggested details showed that, as predicted, the similar-event narrative generated significantly more false recall than the dissimilar-event narrative, $t = 3.0$, $SE = .050$. The rate of false recall of suggested items was significantly greater than the rate of false guesses of control items in both the similar-event condition ($t = 7.19$, $SE = .042$) and the dissimilar-event condition ($t = 5.2$, $SE = .029$).

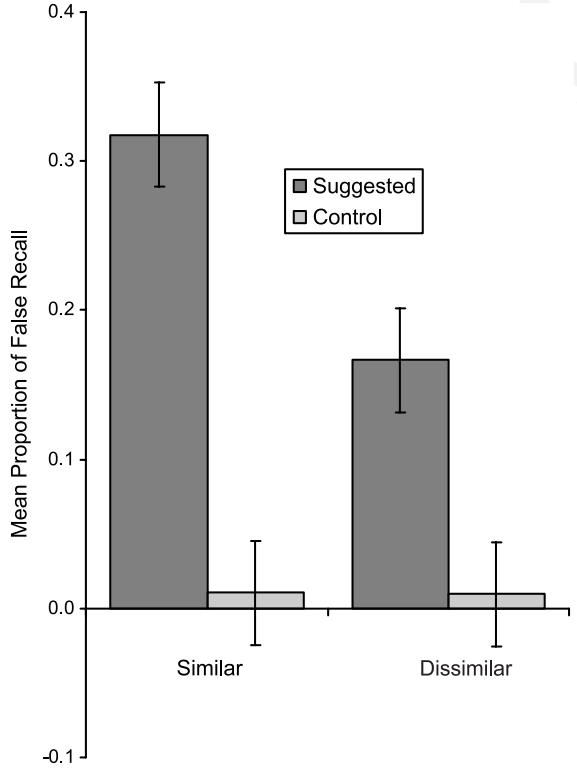

Fig. 3. Mean proportion false recall in Experiment 3 as a function of whether the narrative in which suggestions were imbedded described an event that was thematically similar or dissimilar to the witnessed event. Error bars are 95% confidence intervals calculated using the MSE from the Item Type × Narrative Type ANOVA (see Loftus & Masson, 1994).

Discussion

As predicted, intrusion rates were substantially higher in the similar-event narrative condition than in the dissimilar-event narrative condition. Further discussion of this finding is postponed pending report of Experiment 4.

Experiment 4

In Experiment 2 (in which all parts of the procedure were conducted in a single session), no effect of thematic similarity was obtained. In contrast, in Experiment 3 (in which the narrative and witnessed event were presented in two separate contexts) false reports of suggestions were substantially more common when the witnessed event and narrative were thematically similar than when they were thematically dissimilar. We take this as evidence that disguising the relationship between the two sources creates an opportunity for effects of thematic source-similarity to arise. Of course, this is an across-experiment comparison, and Experiments 2 and 3 differed in several ways in addition to what we take to be the key difference. Experiment 4 was therefore designed as a conceptual replication of Experiments 2 and 3. Half of the participants in Experiment 4 were tested in a procedure that, like Experiment 2, would likely make the relationship between the event and narrative obvious, whereas the remaining participants were tested in a procedure that, like Experiment 3, disguised the relationship between the two sources. We predicted no effect of thematic similarity in the 1-Day (i.e., non-disguised) conditions, but more intrusions from the similar-events narrative than from the dissimilar-events narrative in the 2-Day (disguised) conditions.

Method

Participants

Ninety-six University of Victoria undergraduate students (63% women; age range from 17 to 40 years, mean age = 19.89, $SD = 3.65$) participated in return for optional extra-credit points in an introductory psychology course. There were 24 participants in each condition: 1-Day similar, 1-Day dissimilar, 2-Day similar, and 2-Day dissimilar. Participants were randomly assigned to condition and tested individually.

Materials and procedure

The materials in this experiment were the same as in Experiment 3. Participants were tested individually, half

in a single session and half in two sessions. For participants in the 2-Day condition, the procedure replicated Experiment 3 (except that both experimenters were women). Session 1 was conducted in a basement room and participants were told that they were participating in an imagery study. They were informed that they would hear a narrative twice and then be asked to complete a questionnaire about their imagery for that narrative. After completing these tasks, participants received a bogus debriefing. The following day, participants arrived at a third-floor testing room to take part in an ostensibly unrelated study. The experimenter (a different person from on day one) explained that the study examined how audio soundtracks influence visual perception. Participants watched the critical video and filler video and completed the filler questionnaire and surprise memory test as in Experiment 3. The cued-recall questions were the same as in Experiment 3, and as in the prior experiments participants were asked to base their responses solely on what they had seen in the video and to answer every question (although 27% of participants left one or more questions unanswered). Upon completion of the experiment, participants were debriefed. The same procedure was used in the 1-Day condition, except that one experimenter conducted both sessions in the basement room, with the critical and filler videos shown immediately following the narrative.

Results

Accuracy on filler questions

Similar to Experiments 2 and 3, participants were highly accurate on the three filler questions (mean accuracy = 94%). There were no significant effects of delay or narrative similarity on accuracy (p's > .254), nor was there a reliable interaction ($p = .467$).

Probability of false recall

A preliminary ANOVA included counterbalancing version (A vs. B) as an independent variable, along with item type, similarity, and day. There was a non-significant tendency toward a main effect of counterbalancing narrative, $F(1,88) = 3.48$, $MSe = .043$, $p = .066$, but that factor did not enter into any remotely reliable interactions (all p's > .34) so the analyses below collapse across the counterbalancing factor. Fig. 4 depicts the data.

The proportion of items on which participants reported suggestions, and the proportion of other items on which they guessed non-presented (control) suggestions, were analyzed in a 2 (item type) × 2 (similarity) × 2 (1-Day vs. 2-Day) ANOVA in which the first of these variables was a repeated measure. The key finding was the predicted three-way interaction between these variables, $F(1,92) = 4.10$, $p = .046$, $MSe = .041$, $es = .043$. Separate Item Type × Similarity ANOVAs for the

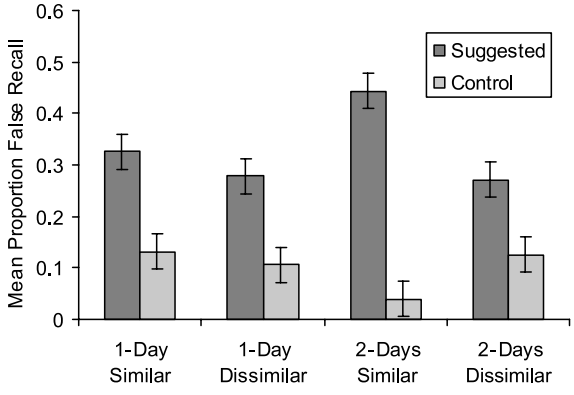

Fig. 4. Mean proportion false recall in the 1-Day and 2-Day conditions of Experiment 4 as a function of whether the narrative in which suggestions were imbedded described an event that was thematically similar or dissimilar to the witnessed event. Error bars are 95% confidence intervals calculated using the MSE from the Item Type × Narrative Type ANOVA (see Loftus & Masson, 1994).

1- and 2-Day conditions showed that there was a reliable effect of item type in each condition (1-Day, $F(1,46) = 22.94$, $p = .000$, $MSe = .035$, $es = .33$; 2-Day, $F(1,46) = 39.18$, $p = .000$, $MSe = .046$, $es = .46$). Moreover, just as predicted, in the 1-Day condition there was no Item type × Similarity interaction ($F < 1$), whereas in the 2-Day condition that interaction was reliable, $F(1,46) = 8.67$, $p = .005$, $MSe = .046$, $es = .16$.

Discussion

As predicted, the 1-Day condition replicated the null effect of thematic similarity obtained in Experiment 2, whereas the 2-Day condition replication the reliable effect of thematic similarity obtained in Experiment 3. Evidently, when event, narrative, and test are all presented as part of the same overall experience, the high degree of concomitance overwhelms any effect of thematic similarity, presumably because participants apprehend the narrative as being related to the event even in the dissimilar-events condition. In contrast, when the relationship between event and narrative is disguised, a thematic-similarity effect emerges.

Why did the interaction take the form of a particularly high rate of false reports in the 2-Day/similar-events condition, rather than a particularly low rate of false reports in the 2-Day/dissimilar-events condition? We speculate that in the 1-Day condition suggestions very often came to mind at test, due to the semantic overlap between questions and suggestions, the recency of suggestions, and the fact that the situation led participants to apprehend the event and narrative as related. It is likely, though, that participants in the 1-Day conditions often correctly attributed those memories to the narrative, and hence did not rely on them as answers

to the questions about the event. Together, these two phenomena led to a relatively low rate of false-reports of suggestions in the 1-Day conditions that did not vary as a function of thematic similarity. In the 2-Day condition, in contrast, thematic similarity played a role in increasing both the likelihood that suggestions would come to mind at test and that they would be accepted as answers, and the 24-h delay (and other aspects of the disguise manipulation) likely lowered the likelihood that participants would spontaneously attribute those memories to the narrative. Consequently, the similar-events narrative produced a high rate of intrusions in the 2-Day condition.

General discussion

The five experiments reported above replicated the core finding of Allen and Lindsay (1998): details mentioned in a narrative description of one event sometimes intruded into participants' reports of a different visual event. These findings substantially increase the applied scope of the eyewitness misinformation effect. Whereas only some witnesses are exposed to misinformation in the context of statements about a witnessed event, virtually all witnesses have past experiences (direct or vicarious) that include details that map onto questions that might be asked in a forensic context. It is worth noting that the false reports in these studies did not merely involve subtle distortions of witnessed details (e.g., misremembering Coke as Pepsi) but rather reports of entire objects that were not witnessed at all.

Consistent with the source-monitoring framework, Experiments 1A, 1B, and 2 found that intrusions were more frequent when the narrative was about the witnessed event than when it was about another event. Presumably, apprehending the narrative as a description of the witnessed event leads participants to think about it during encoding in ways that are similar to how they thought about the witnessed event. In consequence of that high similarity, suggestions are both likely to come to mind and likely to be accepted as memories of the event during test. The null effect of the warning in Experiments 1A and 1B suggests that this is not merely a matter of participants more often knowingly relying on narrative-only memories in the same-event condition.

In contrast to the robust difference in the rate of intrusions from same-event versus different-event narratives, Experiment 2 found no reliable difference in the rate of intrusions from narratives of thematically similar versus dissimilar events. Importantly, however, the intrusion rate was substantially higher in the similar- than dissimilar-events condition of Experiment 3, in which the relationship between the narrative and event was disguised by presenting the two sources in the context of ostensibly separate experiments. Experiment 4 replicated that thematic-similarity effect in the 2-Day condition, and also replicated the null effect of Experiment 2 in the 1-Day condition.

These results, together with findings from prior studies, suggest that whether source-similarity manipulations influence suggestibility effects depends on several factors. First, test instructions must be such that participants are unlikely to report information they know came only from the extra-event source: If participants deliberately use the extra-event source as a basis for answers, source-similarity is unlikely to affect intrusion rates. Second, the test must allow participants to report suggested details (as opposed, e.g., to the McCloskey & Zaragoza, 1985, modified test). Third, the source-similarity manipulation must be sufficiently powerful in the context of other aspects of the task. Some manipulations (e.g., Zaragoza & Lane's, 1994 manipulations of the extent to which participants elaborated and integrated suggestions, and our manipulation of same- vs. different-event narratives) are so powerful that effects can be obtained even when all phases of the experiment are conducted in a single session. Other manipulations (e.g., our manipulation of thematic similarity) may not produce detectable effects when numerous other aspects of the situation highlight the relationship between the event and extra-event information (as in our 1-Day conditions), but yield effects when these other dimensions of background similarity are attenuated to disguise the relationship between event and narrative (as in our 2-Day conditions).

The variation across our experiments, and across previous experiments, in whether source-similarity manipulations affected suggestibility are consistent with the idea that multiple memorial and situational dimensions interact to determine source-monitoring performance. Event memories are complex and multifaceted, and the aspects of memories that are revived and contribute to source attributions vary across situations. Consequently, it is often difficult to predict the effect of manipulating a single dimension of source-relevant information. For example, a conceptual replication of Shaw et al.'s (1997) procedure, with the manipulation of same versus different language during the target event and extra-event information conducted across two ostensibly unrelated sessions to disguise their relationship, might yield a source-similarity effect (i.e., more false reports in the same-languages condition than in the different-languages condition). But it might be that no such effect would be obtained, because people are very skilled at abstracting meanings from the surface form of the language in which they are communicated, and direct tests of memory tend to draw on memory for relatively abstract, conceptual operations (Toth, 2000).

The development of better-specified models of source monitoring depends, in part, on progress toward understanding the crucial but elusive psychological con-

struct of similarity. Even in quite simple situations (e.g., explicit similarity ratings of pairs of pictures or words), perceptions of similarity are complex. One reason for this is that even relatively simple objects, such as two-dimensional geometric shapes, have an indefinitely large number of features or aspects (e.g., angularity, complexity, size, color, and location) and the relative weightings of the contributions of different dimensions to perceptions of similarity varies across contexts, instructions, and individuals (e.g., Medin, Goldstone, & Gentner, 1993). More fundamentally, similarity is not necessarily based on physical features, but can be based on more abstract characteristics such as functions (e.g., hammer and rock) or thematic relations (milk and cow) (e.g., Sloman & Rips, 1998; Wisniewski & Bassok, 1999). As Medin et al. (1993) argued, similarity is not a property of objects, but rather a product of mental processes. Little is known of how mental processes determine degrees of similarity between complex, temporally extended, naturalistic experiences (e.g., witnessing a crime and hearing a story).

Predicting source-monitoring performance is further complicated by the fact that source-monitoring accuracy is not determined solely by the degree of similarity between memories from different sources. Rather, source monitoring is also influenced by the rememberer's goals and orientation at test, decision-making strategies and biases, beliefs and expectations, accessibility, etc. (see Lindsay & Johnson, 2001, pp. 151–152). Thus a complete understanding of source monitoring requires specification of how multiple aspects of memories and situations interact to determine SM accuracy.

The research reported here does not solve these problems, but it does take modest steps in that direction by showing that the effect of a similarity manipulation on intrusion rates depends on contextual variables (i.e., whether the context does or does not lead participants to apprehend two sources as related). Exploring how various aspects and dimensions of past experiences interact with characteristics of the memory testing situation to modulate source-attribution accuracy is a long-term goal for future research (see also Lindsay & Read, 2001). For now, we can conclude that (a) people sometimes intrude details from a narrative of one event into their reports of a different witnessed event, although (b) they do so less often than when the narrative describes the witnessed event itself, and (c) if the relationship between the two different-event sources is disguised at acquisition then across-event intrusions are more likely when the two sources share similar themes.

References

Abeles, P., & Morton, J. (1999). Avoiding misinformation: Reinstating target modality. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 52a*, 581–592.

Abra, J. C. (1972). List differentiation and forgetting. In C. P. Duncan, L. Sechrest, & A. W. Melton (Eds.), *Human memory: Festschrift for Benton J. Underwood* (pp. 25–57). New York: Appleton-Century-Crofts.

Allen, B. P. (1999). Shadows as sources of cues for distance of shadow casting objects. *Perceptual and Motor Skills, 89*, 571–584.

Allen, B. P., & Lindsay, D. S. (1998). Amalgamations of memories: Intrusion of information from one event into reports of another. *Applied Cognitive Psychology, 12*, 277–285.

Birnbaum, M. H. (Ed.). (2000). *Psychological experiments on the Internet*. San Diego: Academic Press.

Bonto, M. A., & Payne, D. G. (1991). Role of environmental context in eyewitness memory. *American Journal of Psychology, 104*, 117–134.

Chambers, K. L., & Zaragoza, M. S. (2002). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition, 29*, 1120–1129.

Chandler, C. C., Gargano, G. J., & Holt, B. C. (2001). Witnessing postevents does not change memory traces, but can affect their retrieval. *Applied Cognitive Psychology, 15*, 3–22.

Eberman, C., & McKelvie, S. J. (2002). Vividness of visual imagery and source memory for audio and text. *Applied Cognitive Psychology, 16*, 87–95.

Frost, P., Ingraham, M., & Wilson, B. (2002). Why misinformation is more likely to be recognized over time: A source monitoring account. *Memory, 10*, 179–185.

Hekkanen, S. T., & McEvoy, C. (2002). False memories and source-monitoring problems: Criterion differences. *Applied Cognitive Psychology, 16*, 73–85.

Hicks, J. L., & Marsh, R. L. (2001). False recognition occurs more frequently during source identification than during old-new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 375.

Holliday, R. E., & Hayes, B. K. (2000). Dissociating automatic and intentional processes in children's eyewitness memory. *Journal of Experimental Child Psychology, 75*, 1–42.

Holliday, R. E., & Hayes, B. K. (2001). Automatic and intentional processes in children's eyewitness suggestibility. *Cognitive Development, 16*, 617–636.

Holliday, R. E., & Hayes, B. K. (2002). Automatic and international processes in children's recognition memory: The reversed misinformation effect. *Applied Cognitive Psychology, 16*, 1–16.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989a). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989b). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118*, 115–125.

Johnson, M. K., Hashtroudi, 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, brain, and belief* (pp. 35–86). Cambridge, MA: Harvard University Press.

Keppel, G., & Underwood, B. J. (1962). Retroactive inhibition of R–S associations. *Journal of Experimental Psychology, 64*, 400–404.

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1077–1083.

Lindsay, D. S., Gonzales, V., & Eso, K. (1995). Aware and unaware uses of memories of postevent suggestions. In M. S. Zaragoza, J. R. Graham, G. C. N. Hall, J. E. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp. 86–108). Thousand Oaks, CA: Sage Publications.

Lindsay, D. S., & Johnson, M. K. (1989a). The eyewitness suggestibility effect and memory for source. *Memory & Cognition, 17*, 349–358.

Lindsay, D. S., & Johnson, M. K. (1989b). The reversed eyewitness suggestibility effect. *Bulletin of the Psychonomic Society, 27*, 111–113.

Lindsay, D. S., & Johnson, M. K. (2001). False memories, fuzzy trace theory, and the source monitoring framework. *Learning and Individual Differences, 12*, 145–161.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*, 297–318.

Lindsay, D. S., & Read, J. D. (2001). The recovered memories controversy: Where do we go from here?. In G. Davies & T. Dalgleish (Eds.), *Recovered memories: Seeking the middle ground* (pp. 71–94). London: Wiley.

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 and Memory, 4*, 19–31.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*, 476–490.

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1137–1151.

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.

McGeoch, J. A. (1936). Retroactive inhibition as a function of the relative amounts of original and interpolated materials. *Psychological Bulletin, 33*, 613–614.

McGeoch, J. A., & McGeoch, G. O. (1937). Studies in retroactive inhibition: X. The influence of similarity of meaning between lists of paired associates. *Journal of Experimental Psychology, 21*, 320–329.

McGovern, J. B. (1964). Extinction of associations in four transfer paradigms. *Psychological Monographs: General and Applied, 78*, 21.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science, 11*, 502–506.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254–278.

Mitchell, K. J., & Zaragoza, M. S. (1996). Repeated exposure to suggestion and false memory: The role of contextual variability. *Journal of Memory and Language, 35*, 246–260.

Mitchell, K. J., & Zaragoza, M. S. (2001). Contextual overlap and eyewitness suggestibility. *Memory & Cognition, 29*, 616–626.

Postman, L., & Stark, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology, 79*, 168–177.

Roediger III, H. L., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language, 35*, 300–318.

Shaw III, J. S., Garcia, L. A., & Robles, B. E. (1997). Cross-language postevent misinformation effects in Spanish–English bilingual witnesses. *Journal of Applied Psychology, 82*, 889–899.

Sloman, S. A., & Rips, L. J. (1998). Similarity as an explanatory construct. *Cognition, 65*, 87–101.

Spencer, R. M., & Weisberg, R. W. (1986). Context-dependent effects on analogical transfer. *Memory & Cognition, 14*, 442–449.

Toth, J. P. (2000). Nonconscious forms of human memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 245–261). London: Oxford University Press.

Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition, 14*, 329–338.

Vecera, S. P., Vogel, E. K., & Woodman, G. F. (2002). Lower region: A new cue for figure-ground assignment. *Journal of Experimental Psychology: General, 131*, 194.

Wisniewski, E. J., & Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cognitive Psychology, 39*, 208–238.

Whittlesea, B. W. A. (2003). On the construction of behavior and subjective experience: The production and evaluation of performance. In J. S. Bowers & C. J. Marsolek (Eds.), *Rethinking implicit memory* (pp. 239–260). London: Oxford University Press.

Zaragoza, M. S., & Koshmider, J. W. (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 246–255.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 934–945.