Mem Cogn (2011) 39:155–170
DOI 10.3758/s13421-010-0005-8

Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing

Ainat Pansky · Einat Tenenboim

Published online: 6 November 2010

Abstract In real-life situations, eyewitnesses often have control over the level of generality in which they choose to report event information. In the present study, we adopted an early-intervention approach to investigate to what extent eyewitness memory may be inoculated against suggestibility, following two different levels of interpolated reporting: verbatim and gist. After viewing a target event, participants responded to interpolated questions that required reporting of target details at either the verbatim or the gist level. After 48 hr, both groups of participants were misled about half of the target details and were finally tested for verbatim memory of all the details. The findings were consistent with our predictions: Whereas verbatim testing was successful in completely inoculating against suggestibility, gist testing did not reduce it whatsoever. These findings are particularly interesting in light of the comparable testing effects found for these two modes of interpolated testing.

Keywords False memory · Suggestibility · Eyewitness memory · Testing effect · Misinformation

Numerous studies have shown that misleading information that is introduced after exposure to an event can have a contaminating effect on subsequent memory reporting of event details (for reviews, see Ayers & Reder, 1998; Belli & Loftus, 1996; Zaragoza, Belli, & Payment, 2007). Following the seminal study by Loftus, Miller, and Burns (1978), a three-stage misinformation paradigm has commonly been used to investigate eyewitness susceptibility to misleading postevent information (MPI). First, the participants are exposed to a sequence of events (e.g., a car stopping at a yield sign), usually by viewing a video or a slide show. Next, they are requested to read a narrative or to answer questions about the previously presented event, in which they are either misinformed about a target item (e.g., the yield sign is referred to as a stop sign; the *misleading* condition), or not (the *control* condition). Finally, the participants are tested for their memory of the original event. Suggestibility is said to occur on the final memory test if the misleading suggestions are more often reported (i.e., recognized or recalled) in the misleading condition than in the control condition.

Replicating the initial findings of Loftus et al. (1978), hundreds of studies employing the misinformation paradigm have shown significant suggestibility effects (for a recent review, see Zaragoza et al., 2007). These findings have raised serious concerns about the trustworthiness of human memory in general and about that of eyewitness testimony in particular, motivating research focused on ways in which the suggestibility effect could be reduced. One attempted tactic that has been found to occasionally reduce (but not eliminate) suggestibility, is (post-) warning misinformed eyewitnesses that they have been exposed to MPI (e.g., Christiaansen & Ochalek, 1983; Eakin, Schreiber, & Sergent-Marshall, 2003; for a review, see Echterhoff, Hirst, & Hussy, 2005). A different approach toward reducing eyewitness suggestibility was to employ an explicit source-monitoring test during final memory testing, in an attempt to encourage rememberers to use more systematic source-monitoring processes during retrieval (e.g., Lindsay & Johnson, 1989). Such attempts have been successful in reducing suggestibility (e.g., Zaragoza & Lane, 1994) and, in some cases, even eliminated the suggestibility effect altogether (e.g., Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989). Nonetheless, most of

This research was supported by The Israel Science Foundation (Grant 898/03-34.2) awarded to Ainat Pansky.

This article is based on the master's thesis submitted by the second author to the Department of Psychology at the University of Haifa.

A. Pansky (✉) · E. Tenenboim
Department of Psychology, University of Haifa,
Haifa 31905, Israel
e-mail: pansky@research.haifa.ac.il

the evidence supports the conclusion that although misled participants are sometimes capable of correctly identifying the source of their suggested memories, they often misidentify them as memories derived from the original event (e.g., Chambers & Zaragoza, 2001; Frost, Ingraham, & Wilson, 2002; Zaragoza & Lane, 1994).

In contrast to a late intervention approach such as a source-monitoring test, by which rememberers scrutinize their responses as (or after) they produce them, an entirely different approach toward reducing suggestibility involves an earlier intervention intended to inoculate one against future suggestibility. This idea can be illustrated by an example of quality control in a manufacturing situation that was put forward by Jacoby, Shimizu, Daniels, and Rhodes (2005). An obvious means of quality control is the inspection and monitoring of manufactured products, in which one rejects those that do not meet the standards. However, quality control may also be accomplished by increasing manufacture constraints so as to decrease the probability of producing defective products to begin with. Clearly, there are advantages for an early intervention, in addition, perhaps, to late interventions such as warnings and source-monitoring tests that were found to be only partially successful.

In the present study, we adopted such an early-intervention approach and attempted to inoculate rememberers against potential contaminating effects of MPI by enhancing the accessibility of event details via neutral memory testing immediately after the event. This approach is based on two separate sets of findings. First, the findings of several studies imply that suggestibility is less pronounced when memory for the target details is stronger or more accessible. For example, Pezdek and Roe (1995) found that children with better memory for an eyewitness event (due to repeated presentation) were less likely to claim to have seen suggested items (see also Henry & Gudjonsson, 2004; Marche, 1999). Recently, Paz-Alonso and Goodman (2008) found greater suggestibility when the MPI was introduced after a delay of 2 weeks than immediately after exposure to the event, attributing this difference to the weakening of event traces over time (see also Loftus et al., 1978). Taken together, these findings suggest that an early intervention that strengthens the memory traces for the target event and/or improves their accessibility can be expected to reduce suggestibility.

The second set of findings on which the present approach is based suggests that immediate memory testing may serve as such an intervention. Consistent with Bjork's (1975) claim that memory testing not only retrieves stored information from memory, but also modifies the state of the memory representations of that information, a remarkable amount of research has demonstrated that taking a memory test improves subsequent retrieval of the tested information, in what is known as the *testing effect* (for recent reviews, see Rajaram & Barber, 2008; Roediger, Agarwal, Kang, & Marsh, 2010; Roediger & Karpicke, 2006a; Roediger, McDermott, & McDaniel, in press). In fact, many laboratory studies have shown that test trials enhance performance on a subsequent test even more than do study trials (e.g., Carrier & Pashler, 1992; Karpicke & Roediger, 2007; Kuo & Hirshman, 1996; Spitzer, 1939; Wenger, Thompson, & Bartling, 1980; Wheeler, Ewers, & Buonanno, 2003). More naturalistic studies simulating eyewitness situations have also demonstrated facilitative effects of testing (e.g., Bornstein, Liebel, & Scarberry, 1998; Poole & White, 1991; Scrivner & Safer, 1988; Warren & Lane, 1995).

Whereas it is well established that testing can inoculate against forgetting, the main question underlying the present study is whether testing can also inoculate against susceptibility to MPI and, if so, under what conditions? A few previous studies have investigated this issue and have yielded conflicting results. Warren and Lane (1995) manipulated the type of initial test (no test, neutral, or misleading) and the type of a second test administered 1 week later (no test, neutral, or misleading). They found that immediate neutral testing was successful in inoculating against both forgetting and suggestibility. They assumed that the initial testing strengthened the memory traces of the event details, consequently making them either more resistant to interference or more accessible than the competing MPI traces. Although not the main focus of their study, Saunders and MacLeod (2002, Experiment 1) similarly found less suggestibility for items that received retrieval practice than for items from the same episode that did not. Whereas the results of these previous studies seem to converge in demonstrating an inoculating effect against suggestibility of immediate interpolated testing, two recent studies have shown the opposite effect (Chan & Langley, in press; Chan, Thomas, & Bulevich, 2009). In these studies, initial cued-recall questions about some event details not only did not inoculate against suggestibility, but actually increased it. This mixed pattern of results across the different studies seems to imply that the effect of initial testing on subsequent suggestibility can be either beneficial or detrimental, and more research is needed in order to elucidate the conditions under which each occurs.

Lane, Mather, Villa, and Morita (2001) theorized that the type of review (i.e., detailed or more general) performed after witnessing an event (and exposure to MPI) can affect suggestibility and found supporting evidence. However, they found no inoculation against suggestibility as a result of either the detailed review or the general review, as compared with no review at all, whether or not the review occurred before or after the exposure to MPI. Note, though, that in contrast to the other studies (as well as our own), Lane et al. used additive, rather than contradictory, MPI.

Thus, it is perhaps not surprising that review of details that were part of the target event did not inoculate against suggestibility involving details that were not.

Somewhat related to the motivation of Lane et al. (2001), in the present study, we examine to what extent inoculation against suggestibility is influenced by the level of responding solicited on the interpolated test regarding each target item. In real-life situations, as opposed to traditional laboratory settings, eyewitnesses often have control over the level of generality at which they choose to report event information. Thus, for example, they may choose to answer a question at a level of generality at which they are likely to be correct (see, e.g., Goldsmith & Koriat, 1999; Pansky, Koriat, & Goldsmith, 2005). According to *fuzzy-trace theory* (FTT; for reviews, see Reyna & Brainerd, 1995; Reyna & Titcomb, 1997; Titcomb & Reyna, 1995), witnesses to an event derive in parallel two types of independent memory representations of each event detail: verbatim representations (i.e., exact surface details of experience) and gist representations (i.e., memory for substance). In accessing these representations, because of the superior memorability and accessibility of gist, especially over time, rememberers tend to choose the highest possible level that complies with task demands, usually favoring gist processing (see Brainerd & Reyna, 2001).

Indeed, several studies have shown that rememberers often choose to respond by reporting gist rather than verbatim information, when given the option to do so (e.g., Goldsmith, Koriat, & Pansky, 2005; Goldsmith, Koriat & Weinberg Eliezer 2002; Pansky, 2010; Pansky & Koriat, 2004; Weber & Brewer, 2008). For example, Goldsmith et al. (2002; see also Goldsmith et al., 2005) have shown that rememberers who are given the option to control the *grain size* of their responses strategically use this option to accommodate the competing goals of accuracy and informativeness. Goldsmith et al. (2005) tested the memory for quantitative information contained in a fictitious eyewitness transcript. When their participants were given control over the grain size of their answers, they chose to provide more coarse-grain answers as the retention interval increased, with the mean percentage of coarse-grain answers increasing from 43% on immediate testing to 61% after 24 hr, and to 75% after 1 week. Using story material, Pansky and Koriat (2004) found that over one third of the target items presented at the subordinate level (e.g., *poodle*) were recalled at the basic level (BL; e.g., *dog*) at immediate testing, exhibiting spontaneous gist, rather than verbatim retrieval. After 1 week, the proportion of items that were presented at the subordinate level and reported at the BL almost doubled. Recently, using a slide show as the target event (the same one as that used in the present study), Pansky (2010) manipulated the timing of the interpolated test and examined the spontaneous level of responding in the absence of specific instructions or examples. The participants were found to respond by reporting gist (i.e., the BL) for about half of the target items when questioned immediately and for about three quarters of the target items when questioned after 48 hr. To summarize, these studies demonstrate that gist retrieval not only is the preferred retrieval mode over time, but also is prevalent even at immediate testing.

Suppose that an eyewitness is questioned to a verbatim specificity on certain event details following a considerable retention interval. On the basis of the vast bulk of findings demonstrating positive effects of interpolated testing for a variety of materials (for reviews, see Roediger et al., 2010; Roediger & Karpicke, 2006a), we would expect such delayed memory reporting of event details to benefit from early interpolated retrieval of the target material. As compared with a no-testing condition, such testing could benefit subsequent memory performance by strengthening the verbatim traces of the original event details, making them more accessible through retrieval practice, or both (see, e.g., Bjork & Bjork, 1992; Brainerd, Reyna, Howe, & Kingma, 1990; Carpenter, 2009; Howe, Courage, & Bryant-Brown, 1993; McDaniel & Masson, 1985; Reyna, 1995). That is, early retrieval of event details at the verbatim level in response to interpolated testing is expected to enhance their accessibility at a later time, yielding a testing effect. Furthermore, this accessibility enhancement is expected to inoculate against MPI (i.e., yield reduced suggestibility) because it is likely to increase the advantage in relative accessibility of the target items, as compared with the competing misleading suggestions on the final test. Suppose, instead, that these event details are retrieved at the gist level. Such interpolated retrieval cannot be expected to directly enhance the accessibility of the corresponding verbatim traces but can be expected to enhance the accessibility of the (tested) gist traces, which, in turn, may serve as effective internal retrieval cues for the subsequent retrieval of the solicited verbatim details. Thus, immediate interpolated retrieval of event details at the gist level may yield an indirect testing effect on subsequent testing of the same details at the verbatim level. However, interpolated gist testing is not expected to yield an inoculation effect against MPI if the presumably more accessible gist traces are consistent with both the original verbatim information and the MPI, as in many misinformation studies, and are therefore inefficient as retrieval cues (see Chandler & Fisher, 1996; Chandler, Gargano, & Holt, 2001).

The present study

In the present study, we used an adaptation of the classic misinformation paradigm in order to test these predictions.

The participants initially viewed a slide sequence depicting a day in a female student's life, containing 16 target details. Subsequently, the participants performed an immediate neutral interpolated memory test containing questions on half of the target items, whereas the other half was assigned to the *untested* condition. For the *gist-testing* group, these questions required gist recall of the target items, whereas for the *verbatim-testing* group they required verbatim recall. Following Brainerd and Reyna (1998), gist and verbatim recall were implemented using two different hierarchical levels: BL and subordinate level,[1] respectively. Thus, participants in the gist-testing group were instructed to provide a one-word response that was expected to represent the target item at the BL (e.g., *chair*), whereas participants in the verbatim-testing group were instructed to provide a two-word response that was expected to represent the target item at the subordinate level (e.g., *wooden chair)*. After 48 hr, all of the participants returned to the lab. They first answered yes/no questions about the slide show, which introduced MPI for half of the target details. Finally, all of the participants were tested for their verbatim memory of the target details at the subordinate level, as in many misinformation studies (e.g., Eakin et al., 2003; Lane & Zaragoza, 2007; Lindsay & Johnson, 1989; Loftus et al., 1978; McCloskey & Zaragoza, 1985). The participants were also asked to indicate, for each of their responses, whether they remembered it from the slide show or were merely guessing. The aim of this procedure was to distinguish between responses accompanied by a phenomenological experience of remembering (i.e., judged as "remember"; see Tulving, 1985) and mere guesses (i.e., judged as "guess") and, thus, to enable an additional examination of the data with guesses excluded. Due to the forced-report format of the final test, it was important to verify that any findings of inoculation against suggestibility or lack thereof apply to cases in which rememberers believed that they were recalling an event item, rather than to mere guesses, allowing a purer investigation of our predictions.

Given that both gist and verbatim testing reinstate the original event by providing retrieval cues that remind one of the original experience (see Howe et al., 1993) and that both entail retrieval practice (see Roediger & Karpicke, 2006a), both modes of interpolated testing were expected to manifest a testing effect for control (i.e., nonmisleading) items. In other words, we predicted a higher proportion of correct responses for tested than for untested items in each of the interpolated-testing groups (gist testing, verbatim testing). Interpolated verbatim retrieval was predicted to further enhance subsequent correct recall of verbatim event details by strengthening the retrieved verbatim representations and/or enhancing their accessibility. This positive effect of verbatim retrieval was also expected to inoculate against suggestibility, such that suggestibility would be less pronounced for verbatim-tested than for untested items. Interpolated gist retrieval, on the other hand, was predicted to contribute indirectly to subsequent correct recall of verbatim responses by enhancing the accessibility of the gist traces, thus increasing the likelihood that they might serve as internal retrieval cues for the solicited verbatim response on the final test. However, no inoculation effect was expected in the gist-testing group, because the practiced gist representations were consistent not only with the correct verbatim responses, but with the incorrect suggested responses as well. Hence, suggestibility on the final memory test was expected to be as pronounced for gist-tested as for untested items.

Method

Participants

The participants were 64 undergraduate students at the University of Haifa, who took part in the experiment for either payment or course credit. Thirty-two participants were randomly assigned to the gist-testing group, whereas the remaining 32 participants were assigned to the verbatim-testing group.

Materials

A 6.5-min slide show was used as the target event. The slide show consisted of still pictures accompanied by a matching soundtrack telling a story about a day in a female student's life. Sixteen concrete items (e.g., *wooden chair*, *mushroom pizza*), each presented visually on a separate slide, constituted the target items (see column 2 in Appendix A for a complete list of the target items).

The interpolated test consisted of eight nonmisleading cued-recall questions about the slide show (see column 5 in Appendix A), each referring to one of the eight target items assigned to the testing condition. All of the questions were presented to the participants in the order in which the target items had appeared in the slide show. The participants in the gist-testing group were directed to provide their responses using only one word that was expected to indicate the BL of the target item. In the verbatim-testing group, for each target item, the same (gist-testing) question was followed by a second question, designed to solicit a

[1] Throughout this article, *subordinate level* refers to a level of specificity that is more detailed than the BL, including both the type of items that are typically used in categorization research (e.g., *kitchen table, poodle*; see Markman & Wisniewski, 1997; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) and the type of items that are typically used in misinformation studies (e.g., *Glamour magazine*, *Folgers coffee*; see McCloskey & Zaragoza, 1985).

response at a more detailed (subordinate) level. Thus, the participant was first asked what a certain item that had appeared in a given scene in the slide show was (e.g., "When Inbal came back home from the pub and started reading her book, what was she sitting on?"). In the second question about that item, the participant was asked what *kind* of that item it was, with the response that she provided in the first phase (e.g., *chair*) inserted by the computer program in the second question (e.g., "What kind of chair?"). An example that was provided in the instructions gave the participants an idea of the level of specificity that was expected for each phase. They also performed one practice trial regarding a nontarget item from the slide show and received feedback. The remaining eight target items were assigned to the untested condition and were not tested at this stage of the experiment.

The misinformation was embedded in a different set of questions about the slide show. In this stage, a total of 16 yes/no questions were presented to the participants: Eight questions included MPI about eight target items (the *misleading* condition) by referring to them using a different subordinate item belonging to the same BL (e.g., "When Inbal came back home from the pub, was the book she was reading while sitting on a plastic chair written in Hebrew?"; see column 4 in Appendix A for a complete list of misleading questions). The other eight target items were not referred to at this stage (the control condition). Instead, eight additional questions, containing no misleading information and not referring to any target item, served as filler questions for all the participants in order to reduce the proportion of misleading questions, thus rendering the introduction of the MPI more subtle. In a further attempt to obscure the misinformation, the participants were instructed to concentrate on the underlined information as the focus of each question (see the example above). The assignment of items to each of the four conditions (untested–control, untested–misleading, tested–control, tested–misleading) was counterbalanced across participants.

The final cued-recall test was identical for all the participants and included 16 questions (one question on each target item; see column 5 in Appendix A). The participants were asked to recall the target items to a verbatim specificity, basing their answers solely on what they had viewed in the slide show, using the same two-question procedure for each item as for the interpolated verbatim-testing phase. An example that was provided in the instructions, as well as another practice trial on a nontarget item, gave the participants an idea of the level of specificity that was expected from them. As in the interpolated test, all the questions were presented in chronological order. The participants were required to answer all of the questions but were asked to indicate for each response whether they were "remembering" or "guessing."

Procedure

In the first stage of the experiment, the participants viewed the slide show. After completing a nonverbal filler task of solving Raven's progressive matrices for approximately 10 min, the participants proceeded to the second stage of the experiment, in which they were given the interpolated cued-recall test, containing eight questions that required either gist or verbatim reporting of the target items, depending on the group of interpolated-testing mode to which the participant was assigned. Following 48 hr, the participants returned to the lab for the remaining stages. In the third stage, the participants were asked to answer yes/no questions about the slide show, half of which introduced misinformation. Subsequently, participants engaged in an additional 10-min nonverbal filler task of number series and were finally requested to answer the final cued-recall test for the fourth and final stage of the experiment. See Appendix B for the instructions that were presented to the participants at the various stages of the experiment.

The experimental design was a 2 (interpolated-testing mode: gist-testing, verbatim-testing) × 2 (interpolated-testing condition: untested, tested) × 2 (misinformation condition: control, misleading) mixed factorial design, with interpolated-testing mode manipulated between subjects and interpolated-testing condition and misinformation condition manipulated within subjects.

Results

Two independent judges determined, for each response that the participants provided on the final cued-recall test, whether it was identical to the event item (i.e., correct), identical to the suggested item (i.e., suggested), or neither. Table 1 presents the recall proportions of each response type in each experimental condition. The two judges also reviewed the participants' responses on the interpolated test and determined whether each response was (1) correct or incorrect and (2) specified at the BL or at the subordinate level. The separate classifications made by these two judges yielded 91.3% agreement, which ascended to 97% after a mutual discussion. A third and final judge determined the scoring of the controversial 3% of the responses.

Before turning to the main analyses, the participants' responses on the interpolated test were examined in order to determine to what extent the participants in each testing group complied with our instructions. For the gist-testing group, we found that 85.2% of the responses on the interpolated test were provided, as intended, at the BL, whereas the remaining 14.8% were provided at the subordinate level (which can be reported, at least in some cases, using a single word, e.g., *Subaru*, *Danone*). These

Table 1 Recall proportions on the final memory test of correct target items, suggested items, and other intrusions, as a function of interpolated-testing mode, interpolated-testing condition, and misinformation condition

Interpolated-Testing Mode	Interpolated-Testing Condition	Misinformation Condition	Type of Response					
			Correct		Suggested		Other Intrusions	
			M	SD	M	SD	M	SD
Gist Testing	Untested	Control	.27	.25	.11	.15	.62	.27
		Misleading	.20	.18	.43	.26	.37	.21
	Tested	Control	.47	.28	.10	.15	.43	.26
		Misleading	.36	.25	.36	.24	.28	.24
Verbatim Testing	Untested	Control	.27	.22	.06	.11	.67	.23
		Misleading	.18	.23	.38	.26	.44	.27
	Tested	Control	.46	.25	.11	.14	.43	.26
		Misleading	.46	.28	.16	.16	.38	.24

percentages imply that the gist-testing manipulation was mostly, but not entirely, effective. Examination of the data of the verbatim-testing group revealed that 94.5% of all interpolated responses were provided at the subordinate level, as intended, whereas the remaining 5.5% were provided at the BL. Thus, the large majority of the responses provided by this group seem to indicate compliance with the instructions. Nevertheless, we repeated the main data analyses while excluding cases in which the participants provided inappropriate-level responses on the interpolated test and found the same pattern of results as those obtained for all of the responses. Therefore, we report only the analyses including the full set of responses.

In addition, we sought to ensure that performance in the (baseline) untested condition was equivalent in the two interpolated-testing modes, so that any differences found between them could be attributed to the type of interpolated retrieval. To this end, two sets of analyses were conducted on the previously untested items from the final memory test. First, the proportion of correct responses was analyzed using a mixed-model ANOVA with interpolated-testing mode (gist testing, verbatim testing) as a between-subjects factor and misinformation condition (control, misleading) as a within-subjects factor. Performance was found to be comparable for the two interpolated-testing groups, since neither the main effect of interpolated-testing mode nor its interaction with misinformation condition was significant, $F<1$ for both. The same analysis was performed on the proportion of suggested items recalled, revealing once again a nonsignificant effect of interpolated-testing mode, $F(1, 62)=2.093$, $MSE=.039$, *ns*, $\eta^2=.033$, and a nonsignificant interaction between interpolated testing mode and misinformation condition, $F<1$. Thus, it seems safe to conclude that performance on the final test was equivalent for the untested items in the two interpolated-testing groups, providing an equated baseline against which the effect of interpolated-testing mode could be examined. Moreover, these analyses indicate that the untested items of both interpolated-testing groups were not influenced by the different retrieval modes of the tested items. Hence, this finding can be taken as evidence that our target items were processed independently of one another, as intended, and that the processing mode of the tested items did not carry over to the untested items.

In the second set of analyses, we examined to what extent performance in the two interpolated-testing mode conditions was equivalent in the interpolated-testing stage. Obviously, a simple comparison between the accuracy of the responses on the interpolated test for the two interpolated-testing modes is inappropriate, given that accurately responding at the BL, which was anticipated for the gist-testing group, is a less difficult task than accurately responding at the subordinate level, which was anticipated for the verbatim-testing group. Indeed, whereas 79% of the interpolated responses provided by the gist-testing group were accurate, only 49% of the interpolated responses provided by the verbatim-testing group were accurate. However, a comparison of correct BL recall in the two interpolated-testing mode groups, which seems more appropriate, revealed that the BL was correctly recalled in 77% of the interpolated responses of the verbatim-testing group, a percentage that is not significantly different from that for the gist-testing group (79%), $t(62)=0.555$, *ns*, $d=0.05$. This comparison confirms that performance in the two interpolated-testing mode groups was comparable, not only in terms of the (baseline) untested condition on the final test, but also in terms of the accuracy of the BL provided for tested items on the interpolated test. Therefore, we conclude that any differential testing or inoculation effects found between the two interpolated-testing groups can be attributed to the different testing modes, rather than to a priori sampling differences.

For the main analyses, two sets of analyses were performed, each on a different dependent variable: (1) An examination of potential testing effects was performed on the proportion of correct recall of the nonmisleading target items, and (2) an examination of potential inoculation effects against MPI was performed on the proportion of suggested items recalled.

Testing effects

Testing effects were examined in the control (i.e., nonmisleading) conditions of the two interpolated-testing modes. We predicted that, as compared with the untested conditions, both verbatim and gist testing would yield testing effects. In this set of analyses, interpolated-testing mode (gist testing, verbatim testing) served as a between-subjects factor, whereas interpolated-testing condition (untested, tested) served as a within-subjects factor. A mixed-model ANOVA for repeated measures revealed an overall testing effect, $F(1, 62)=28.514$, $MSE=.043$, $p<.0001$, $\eta^2=.315$, with a higher proportion of accurate responses following interpolated testing (.47) than following no interpolated testing (.27). The testing effect was comparable for the two interpolated-testing modes, and, consequently, the interaction between interpolated-testing mode and interpolated-testing condition was not significant, $F<1$. As predicted, the testing effect was significant for both gist testing and verbatim testing, with a higher proportion of correct responses on the final memory test for gist-tested items (.47) than for untested items (.27), $t(31)=3.498$, $p<.001$, $d=0.62$, and for verbatim-tested items (.46) than for untested items (.27), $t(31)=4.132$, $p<.0001$, $d=0.73$. Thus, correct recall of the target items on the final memory test was found to benefit from interpolated questioning regardless of the interpolated-testing mode that was solicited (i.e., verbatim or gist).

Inoculation against suggestibility

As was previously mentioned, the suggestibility effect is operationally defined as the difference between misleading and control conditions in the proportion of suggested items recalled. This comparison assesses the extent to which the suggested items are more likely to be recalled after they have been suggested than they are likely to be recalled spontaneously, without having been suggested. We predicted that verbatim testing, but not gist testing, would inoculate against suggestibility, such that suggestibility on the final memory test would be less pronounced in the verbatim-testing condition than in the *untested* condition but would not be less pronounced in the gist-testing condition than in the untested condition. In this set of analyses, interpolated-testing mode (gist testing, verbatim testing) served as a between-subjects factor, whereas interpolated-testing condition (untested, tested) and misinformation condition (control, misleading) served as within-subjects factors.

As was expected, the ANOVA revealed a significant three-way interaction between interpolated-testing mode, interpolated-testing condition, and misinformation condition, $F(1, 62)=4.357$, $MSE=.035$, $p<.05$, $\eta^2=.066$. For the gist-testing group, a higher proportion of suggested items was recalled in the misleading (.40) than in the control (.11) condition, resulting in a significant suggestibility effect, $F(1, 31)=51.566$, $MSE=.052$, $p<.0001$, $\eta^2=.625$. Importantly, the interaction between misinformation condition and interpolated-testing mode was not significant, $F<1$, with a comparable suggestibility effect found for gist-tested items (mean difference between misleading and control conditions = .26), $t(31)=5.075$, $p<.0001$, $d=0.90$, as for untested items (mean difference between misleading and control conditions = .32), $t(31)=5.680$, $p<.0001$, $d=1.00$. For the verbatim-testing group, an overall suggestibility effect was found as well, $F(1, 31)=28.169$, $MSE=.038$, $p<.0001$, $\eta^2=.476$, with participants recalling significantly more suggested items on the final test in the misleading (.27) than in the control (.09) condition. However, as was expected, in contrast to the gist-testing group, here the interaction between interpolated-testing condition and misinformation condition was significant, $F(1, 31)=17.890$, $MSE=.030$, $p<.0001$, $\eta^2 = .366$. In the untested condition, we found a significant suggestibility effect, $t(31)=6.225$, $p<.0001$, $d=1.10$, with participants recalling more suggested details in the misleading (.38) than in the control (.06) condition. In support of our prediction, in the verbatim-testing condition, we found no significant difference between the misleading (.16) and control (.11) conditions, $t(31)=1.315$, *ns*, $d=0.23$, exhibiting no suggestibility effect.

As was previously stated, remember/guess judgments were solicited on the final memory test in an attempt to distinguish between recollections and mere guesses. Table 2 presents the proportion of "remember" responses in each experimental condition. Each of the analyses performed on the entire set of responses was also performed on the subset of "remember" responses (excluding those classified as guesses), and yielded the same pattern of results.[2] Overall, a higher proportion of suggested items was recalled in the misleading (.30) than in the control (.06) condition, resulting in a significant suggestibility effect, $F(1, 27)=$

[2] Whereas all 64 participants were included in the analyses performed on the entire set of data, fewer participants were included in the analyses performed on "remember" responses only, because some of them had classified all their responses in at least one condition as guesses.

Table 2 Mean proportion of 'remember' responses for *tested* and *untested* items, as a function of interpolated-testing mode and misinformation condition

Interpolated-Testing Mode	Misinformation Condition	Untested		Tested		Total	
		M	SD	M	SD	M	SD
Gist Testing	Control	.44	.25	.73	.21	.58	.27
	Misleading	.64	.28	.80	.25	.72	.28
	Total	**.54**	**.29**	**.76**	**.23**	**.65**	**.28**
Verbatim Testing	Control	.43	.28	.74	.23	.59	.30
	Misleading	.63	.23	.83	.18	.73	.23
	Total	**.53**	**.27**	**.79**	**.21**	**.66**	**.28**

49.470, MSE=.034, p<.0001, η^2=.647. As can be seen in Fig. 1, the suggestibility effect in the gist-testing condition (mean difference between misleading and control conditions = .26), $t(27)$=4.816, p<.0001, d=0.91, and the untested condition (mean difference between misleading and control conditions = .23), $t(27)$=5.227, p<.0001, d=0.99, was comparable, resulting in a nonsignificant interaction between misinformation condition and interpolated-testing mode, F<1. Thus, as was predicted, interpolated gist testing did not inoculate against suggestibility on the final test. In contrast, for the verbatim-testing group, a suggestibility effect was found for untested items (mean difference between misleading and control conditions = .21), $t(26)$=4.665, p<.0001, d=0.90, but not for verbatim-tested items (mean difference between misleading and control conditions= .01), $t(26)$=0.273, *ns*, d=0.05, with a significant interaction between interpolated-testing mode and misinformation condition, $F(1, 26)$=20.800, MSE=.016, p<.0001, η^2=.444. Evidently, interpolated memory testing that required reporting of verbatim responses completely eliminated the suggestibility effect, thus inoculating rememberers against the contaminating influence of MPI.

Discussion

In the present study, we investigated the conditions under which memory for a target event may be inoculated against the contaminating effect of MPI, via immediate interpolated memory testing. We also examined the effects of gist and verbatim interpolated testing on subsequent correct recall of nonmisled event details. In what follows, we will begin by briefly discussing each set of findings separately and will then proceed to a discussion of the overall pattern of results.

Inoculation effects against MPI

In support of our predictions, we found an inoculating effect for interpolated verbatim testing, but not for interpolated gist testing, such that verbatim testing eliminated the suggestibility effect, whereas gist testing failed to reduce it whatsoever. This pattern of results was obtained not only for the entire set of responses on the final test, but also when guesses were excluded. These results are consistent with those of Warren and Lane (1995), who found that an immediate neutral memory test attenuated the

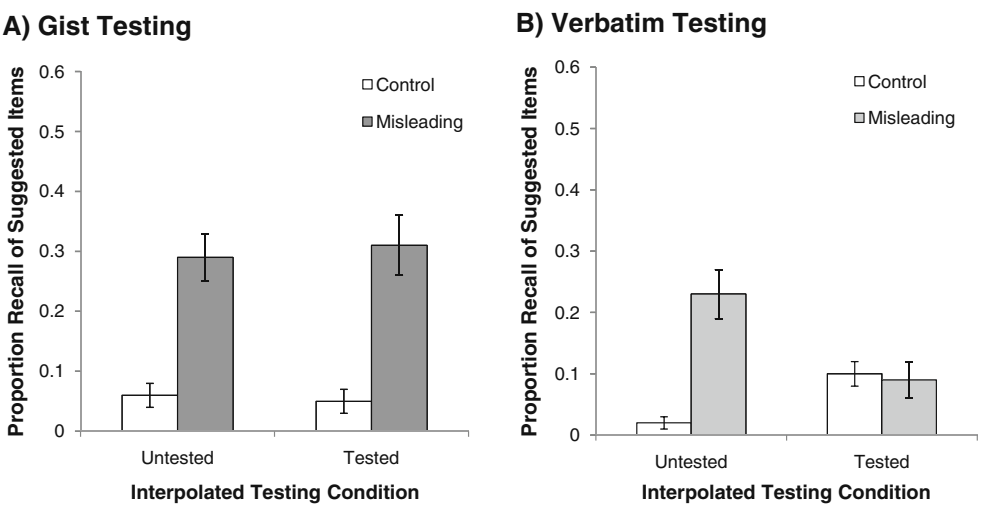

Fig. 1 Mean proportion of suggested items recalled on the final memory test in the gist-testing (panel **A**) and verbatim-testing (panel **B**) groups, as a function of interpolated-testing condition and misinformation condition, data excluding responses classified as guesses. Error bars indicate ±1 *SEM*

harmful effects of later suggestion. Although their experimental materials are not specified in detail, they nonetheless obtained clear evidence of a reduced suggestibility effect for an immediately tested group, in comparison with a group that had not previously been tested, such as we found for the verbatim-testing group. Our results are also consistent with those of Saunders and MacLeod (2002), who found less suggestibility for items that received retrieval practice than for items from the same episode that did not.

Our findings differ from those of Lane et al. (2001, Experiment 2), who found no inoculation against suggestibility for either participants who reviewed the main points of the event or participants who reviewed the event in detail, as compared with participants who did not review the event at all. However, in contrast to our study, Lane et al.'s study involved additive, rather than contradictory, MPI, a critical difference that may explain the differential findings. Chan et al. (2009) and Chan and Langley (in press) found increased suggestibility as a result of immediate interpolated testing, a finding that is highly inconsistent with ours. However, in contrast to our experiment, in these studies the target event was much longer and richer in detail; interpolated testing (i.e., whether or not it took place) was manipulated between subjects; misinformation was introduced via an audio narrative; and the retention intervals were different from ours. Given the numerous differences between these studies and ours, we cannot determine which of them accounts for the different findings. Clearly, the mixed pattern of results across the different studies calls for future research that will shed more light on the factors that affect the extent to which prior testing is beneficial (or detrimental) in terms of suggestibility.

Testing effects

The second set of effects that we examined in this study concerns performance on the control items—that is, items for which MPI was not introduced. On the basis of the extensive literature on the testing effect in which a variety of different materials were used (for recent reviews, see Roediger et al., 2010; Roediger & Karpicke, 2006a; Roediger et al., in press), superior memory performance was expected for previously tested than for untested items. Indeed, an examination of the control data of the two interpolated-testing modes revealed a testing effect for both interpolated gist and verbatim testing, as evident in a higher proportion of correct responses for tested than for untested items. Hence, an immediate cued-recall test that required either gist or verbatim responses on selected items benefited the later recall of these items, provided that no misleading information about the target items was introduced between interpolated and final testing (see the preceding section).

These findings are consistent with many previous findings of testing effects obtained using complex materials such as a filmed event or prose, immediate interpolated testing, and a delayed final memory test (e.g., Chan & Langley, in press; Chan, McDermott, & Roediger, 2006; LaPorte & Voss, 1975; Poole & White, 1991; Roediger & Karpicke, 2006b; Spitzer, 1939; see also Goodman & Quas, 2008; Roediger et al., 2010), as we employed in our study.

Discussion of the overall pattern of results

The effects we obtained following interpolated verbatim testing, the prevalent mode of testing that was used in previous studies, are consistent with previous findings: As compared with untested items, items that were tested to a verbatim level shortly after the witnessed event were more likely to be correctly recalled on a delayed memory test (e.g., Chan et al., 2006; Poole & White, 1991; Warren & Lane, 1995) and were more resistant to misleading information (e.g., Saunders & MacLeod, 2002; Warren & Lane, 1995). The testing effect can be attributed to strengthening of the memory trace, retrieval practice, or both (see, e.g., Bjork & Bjork, 1992; Brainerd et al., 1990; Carpenter, 2009; Howe et al., 1993; McDaniel & Masson, 1985; Reyna, 1995). The inoculation effect against suggestibility can be attributed to a greater resistance to the MPI following interpolated testing (see, e.g., Hall, Loftus, & Tousignant, 1984; Loftus, 2005; Tousignant, Hall, & Loftus, 1986), the enhanced accessibility of an earlier retrieved item, as compared with that of the competing misleading item on the final test (see, e.g., Brainerd & Reyna, 1998; Reyna & Titcomb, 1997; Titcomb & Reyna, 1995), or both.

However, the main innovation of the present study relates to the seemingly puzzling effect of interpolated testing that entails gist retrieval, a retrieval mode that has been found to be quite frequent in free-report situations (see, e.g., Goldsmith et al., 2005; Pansky, 2010; Pansky & Koriat, 2004). On the one hand, interpolated gist questioning was found to contribute considerably to the correct recall of target details in the control condition (i.e., a testing effect), but on the other hand, it did not reduce the false recall of suggested details in the misleading condition whatsoever. We suggest that the key to resolving this puzzle is mainly rooted in understanding the role of the previously retrieved gist representations as retrieval cues on the final memory test (for reviews, see Chandler & Fisher, 1996; Rajaram & Barber, 2008). With regard to nonmisleading (i.e., control) items, we propose that the interpolated gist testing increased the accessibility of the retrieved BL representations, which subsequently served as effective (internal) retrieval cues for the corresponding subordinate items on the final memory test, thus giving them an

advantage over the untested items. An additional analysis that we conducted on the control data supports this possibility. This analysis compared the gist-tested and the untested items with regard to the accuracy of the BL responses (i.e., the responses to the first question concerning the target item) on the final memory test. Indeed, participants correctly recalled the BL of gist-tested items (.78) more frequently than that of untested items (.63), $t(31)=2.552$, $p<.01$, $d=0.45$. Presumably, this enhanced accessibility of the BL representations on the delayed test as a result of interpolated gist testing rendered them more likely to serve as effective retrieval cues for the solicited verbatim details, thus yielding a testing effect. By contrast, in the misleading condition, the gist representations that became more accessible as a result of interpolated testing were not likely to constitute effective retrieval cues on the final memory test, because they served as retrieval cues for both the target details and the more recent suggested details, since they shared the same BL or gist (see Chandler & Fisher, 1996; Chandler et al., 2001). Consequently, interpolated gist testing yielded neither a reduction nor an elimination of the suggestibility effect. In other words, according to this account, rememberers utilized their previously retrieved BL representations as internal retrieval cues for recalling the target subordinate items on the final memory test, a strategy that did not prove to be very useful in the face of BL-consistent MPI.

Although we believe that this retrieval cue account best explains the overall pattern of our results, a potential artifact might have also played a role. It could be that interpolated gist questioning induced the participants to covertly retrieve their answers at the subordinate level despite overtly reporting them at the BL as instructed, thus enhancing the accessibility of their verbatim representations, as compared with that of the untested items. Indeed, examination of the accuracy of the interpolated verbatim responses in the verbatim-testing group showed that 49% of the responses were correct, indicating that verbatim memories were quite accessible at the time of the interpolated retrieval. Thus, it is possible that an interpolated question that required only reporting of the BL (e.g., *chair*) may have involved covert access to the corresponding subordinate representation (e.g., *wooden chair*) of these items. Note that such verbatim retrieval, if it occurred, was covert in large part, given that in this condition the participants were instructed to respond to the questions using one word only and that they indeed complied with these instructions in the majority of the cases (85.2%). If covert verbatim retrieval alone underlay the testing effect obtained in the gist-testing group, we would have anticipated an inoculating effect against MPI, but no such effect was found. Nonetheless, it is possible that covert verbatim retrieval may have operated in parallel with the enhancement of BL accessibility in yielding the null inoculation effect that we found for the gist-testing group. Thus, subordinate items that were highly accessible at the time of the interpolated gist testing may have exhibited covert verbatim retrieval, manifesting reduced suggestibility on the final test for these items. In contrast, items retrieved only at the BL on the interpolated test may have been more likely than untested items to later serve as "effective" retrieval cues on the final test for the recently presented misleading items (that were consistent with that BL), resulting in enhanced suggestibility for these items. Therefore, the observed absence of an inoculating effect for retrieved gist items could have possibly resulted from these two opposing trends: on the one hand, reduced suggestibility for items that were covertly retrieved at the verbatim level, and, on the other hand, enhanced suggestibility for items for which the practiced BL representations served as ineffective retrieval cues.

Finally, an apriori advantage of interpolated verbatim testing over interpolated gist testing in the present study should be considered with regard to our findings. According to the transfer-appropriate processing framework (see Morris, Bransford, & Franks, 1977), greater similarity between encoding and retrieval processes yields superior memory performance. Subsequent studies have applied this approach to interpolated testing, examining whether a stronger match between the processes engaged during the interpolated and final tests results in better performance on the final test (see, e.g., Carpenter & DeLosh, 2006; Johnson & Mayer 2009; Kang, McDermott, & Roediger, 2007). Whereas earlier studies have shown evidence that a match between the format of the interpolated and final tests yields stronger transfer effects than does a mismatch (Duchastel & Nungester, 1982; McDaniel, Kowitz, & Dunay, 1989), more recent studies have not replicated these findings (e.g., Carpenter & DeLosh, 2006; Kang et al., 2007). In the present study, because the final memory test was of a verbatim format that was identical to the interpolated verbatim, but not to the interpolated gist, test format, one might argue that this could, at least partially, explain the inoculating effect found for verbatim-tested, but not for the gist-tested, items. Yet the comparable testing effects that were found for the verbatim-testing and the gist-testing groups speak against this possibility. Naturally, had the degree of correspondence between our interpolated and final test formats contributed to the inoculating effect against suggestibility found for verbatim tested items, we would have expected to also obtain a larger testing effect for the verbatim-testing than for the gist-testing group. Given that the two testing effects were comparable, it appears that transfer-appropriate processing did not play a significant role in inoculating against suggestibility found for our verbatim-testing (but not for our gist-testing) group.

Implications and conclusions

In the present study, both immediate verbatim and immediate gist testing were found to enhance the likelihood of correct recall of target items on a delayed cued-recall test. However, whereas immediate gist testing did not reduce suggestibility, immediate verbatim testing was successful in inoculating against suggestibility. Thus, in addition to late interventions, such as postwarnings about the MPI (e.g., Christiaansen & Ochalek, 1983; Eakin et al., 2003) and source-monitoring tests (Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989; Zaragoza & Lane, 1994), which have been shown to be partially successful in alleviating the contaminating effect of MPI, the present findings demonstrate the success of an early-intervention approach via immediate cued recall of event details, requiring verbatim precision, in eliminating suggestibility altogether.

There are conceptual similarities between our findings and earlier findings obtained by Henkel (2004), using a different paradigm that did not involve the introduction of MPI. She found that source misattributions (by which imagined items were claimed to have been viewed) were less frequent for participants who had taken an earlier source-monitoring test than for those who were merely asked to recall the items earlier (without specifying their source). Both our findings and hers seem to imply that interpolated tests that are less demanding than the subsequent criterial test (e.g., gist recall/recall without source discrimination, as opposed to verbatim recall/source discrimination) might be less beneficial than interpolated tests that are as demanding as the criterial test, at least with regard to certain cases of suggestibility and source confusions on the criterial test.

A large body of research suggests that the use of open-ended (free recall) neutral questioning that allows one to report only information that one feels sure about and at a level of detail of one's own choosing, enhances memory accuracy considerably, as compared with the use of more focused questions (e.g., Fisher, 1995; Fisher, Falkner, Trevisan, & McCauley, 2000; Powell, Fisher, Wright, Brewer, & Williams, 2005; Wells et al., 2000). However, other studies have shown that rememberers often spontaneously respond to such questions by retrieving gist representations (Brainerd & Reyna, 2002; Pansky, 2010; Pansky & Koriat, 2004). In the present study, responding to open-ended questions by retrieving gist representations of the target details did not seem to be effective in inoculating against subsequent exposure to misleading information about these details. Thus, whereas open-ended questioning may be optimal in terms of memory accuracy on a given interview, it is important to investigate to what extent and under which conditions it is also optimal, in the long run, for memory accuracy on subsequent tests, particularly when detailed information is then required. Perhaps an open-ended questioning technique such as the cognitive interview (Fisher & Geiselman, 1992), which has been shown consistently to produce *more detailed* reports from eyewitnesses without reducing accuracy, can also be effective in counteracting the effects of subsequent introduction of misinformation (see Memon, Zaragoza, Clifford, & Kidd, 2010, for recent findings in this direction).

In conclusion, the present study demonstrates that the mode of interpolated testing is an important factor that should be taken into consideration with regard to the potentially contaminating effects of MPI. Directing witnesses to provide detailed verbatim responses during immediate questioning can, at least under some conditions, minimize or even eliminate future unforeseeable interferences.

Appendix A

Table 3 Target items, misleading items, misleading post-event questions and test questions used in this study (translated from Hebrew)

Item number	Target item	Misleading item	Misleading question	Test question
1	*Ma'ariv* newspaper	*Yediot-Acharonot* newspaper	Were the pajamas that Inbal was wearing while reading *Yediot-Acharonot* newspaper black? (no)	What was the reading material that was placed on Inbal's bed when she called her friend Yael to invite her for a get together?
2	*Meridol* toothpaste	*Elmex* toothpaste	Were there tweezers on the bathroom shelf, where the *Elmex* toothpaste was placed? (yes)	When Inbal washed her face before she went out to meet Yael, what was in the tube that was placed on the bathroom's shelf?
3	Red grapefruit	Yellow grapefruit	While Inbal was preparing herself a drink in the morning and the phone was ringing, was the plate containing the yellow grapefruit placed on the microwave? (no)	Which fruit was placed on the kitchen counter, while Inbal was preparing herself a drink in the morning and the phone was ringing?

Table 3 (continued)

Item number	Target item	Misleading item	Misleading question	Test question
4	Black pen	Blue pen	Did the notepaper on which Inbal wrote the message for her father using the blue pen have a square shape? (yes)	Which writing utensil did Inbal use to write her father the message delivered to him over the phone?
5	Orbit gum	*Must* gum	When Danny was about to take his sister Inbal to her friend Yael, and when Inbal found some *Must* gum in her pocket, was the front door closed? (no)	Which sweet did Inbal take out of her pocket, when she was standing outside the door of her house waiting for her brother Danny to give her a lift to her friend Yael?
6	(the letter) S	(the letter) B	While Inbal sat on the sofa across from her friend Yael and reminisced about the time they served in the army, did the cup on which B was printed contain tea? (yes)	Which symbol was printed on the cup Inbal was holding in her right hand, when she was sitting across from Yael and the two reminisced about the time they served in the army?
7	Poodle dog	Cocker Spaniel dog	When Itai showed up on the other side of the street holding a Cocker Spaniel dog, did he pass by a big dumpster? (no)	Which animal did Itai hold in his arms when he met Inbal on her way back home from her visit at Yael's?
8	*Shufersal* supermarket	*Hypercol* supermarket	Were there shopping carts at the entrance of the *Hypercol* supermarket? (yes)	In which store did Inbal's mother realize she forgot her purse?
9	Subaru car	Suzuki car	Did Inbal's mother hold an umbrella when she took the grocery bags out of the trunk of the Suzuki car? (no)	What was the vehicle out of which Inbal's mother took grocery bags, before she entered the house?
10	Danone yogurt	Yoplait yogurt	Did Inbal's father hold a teaspoon in his hand while he was eating the Yoplait yogurt and watching TV? (yes)	What kind of dairy product did Inbal's father eat while watching TV?
11	Crisphead lettuce	Romaine lettuce	Was the salad bowl in which Inbal placed the pieces of the romaine lettuce white? (no)	Which vegetable was placed in a salad bowl on the kitchen's table, when Inbal's mother came home and the two discussed what Inbal should wear when she went out with Itai that evening?
12	Mushroom pizza	Olive pizza	Was there wire rack in the oven in which Inbal spotted an olive pizza? (yes)	What was in the oven when Inbal was trying to identify the source of the delicious scent and accidentally burned her finger?
13	*Neka 7* shampoo	*Hawaii* shampoo	Was the song that Inbal was singing in the shower when she reached out for the *Hawaii* shampoo, a song of *Haim Moshe*? (no)	Which personal care product did Inbal reach for when she was in the shower, singing to herself?
14	Carlsberg beer	Heineken beer	When Itai and Inbal sat at the pub, was the candle, which was placed between the Heineken beer and the glass of wine, lit? (yes)	Which alcoholic beverage, other than wine, did Inbal and Itai order when they sat across from each other at the pub?
15	Silver ring	Gold ring	Did the drawer, which contained a book, a box with a gold ring in it, and a purple booklet, have a colorful handle? (no)	When Inbal opened the drawer in order to find a book that she wanted to read, which piece of jewelry was placed inside the drawer in the open box that was near the book?
16	Wooden chair	Plastic chair	When Inbal came back home from the pub, was the book she was reading while sitting on the plastic chair written in Hebrew? (yes)	When Inbal came back home from the pub and started reading her book, what was she sitting on?

All words that appear in *italics* refer to well-known Israeli products, brands, or celebrities.

Appendix B

Instructions for the witnessed/slide show event (Session 1)

You will now be presented with a slide show. Please concentrate and attend to it carefully. We will later ask you several questions about the slide show.

Instructions for the interpolated test (Session 1)

In this section, we are interested in examining your memory regarding some items that had appeared in the slide show that you viewed at the beginning of this session. For this purpose, we will ask you a few questions.

Verbatim-testing version	Gist-testing version*
You will be asked two questions about each item. In the first question, you will be asked **what** was the item that had appeared in a certain scene. The question refers to the underlined word. For example: "What did Inbal drink at the neighborhood pub?" The correct answer is "wine". In the second question, we will ask you to provide a more detailed answer regarding the item you provided in response to the first question: we will ask you **what kind** of item was it. For example, if your answer to the first question was "wine", you will now be asked "What kind of wine?" The correct answer is "red ".	Each question refers to a specific item, which you are asked to provide **using a single word**. The question refers to the underlined word. For example: "As she was getting ready to leave the house for her date with Itai, what object did Inbal put into her handbag?" The correct answer is "wallet".

*Each participant received one of these two versions of instructions, depending on the interpolated-testing group to which he/she was assigned.

When answering the questions, please do your best and try to avoid answers such as "I don't remember" or "I don't know". If you don't remember a certain item, give the most suitable answer that you can. Consider each response carefully, because as soon as you press the OK button, you will not be able to change your response. Once you have finished reading these instructions, please raise your hand and call the experimenter.

Instructions for the yes/no questions (Session 2)

In this section, you will be asked to answer a series of yes/no questions regarding the slide show that you viewed in the previous experimental session. Please answer "yes" or "no" to each question, with reference to the underlined word.

Examples: 1) "Was the laundry hamper in Inbal's house made of wicker?" The correct answer is "yes". 2) "Was the message that Inbal wrote for her father about a call from his cousin?" The correct answer is "no", because it was her father's friend who had called.

Please read each question carefully and choose the best answer in your opinion. Note that after you answer a question, you will not be able to change your response.

Instructions for the final memory test (Session 2)

In this section, we aim to examine once again your memory regarding some items that had appeared in the slide show that you viewed in the first experimental session.

You will be asked three questions about each item. In the first question, you will be asked **what** was the item that had appeared in a certain scene. The question refers to the underlined word. For example: "What did Inbal drink at the neighborhood pub?" The correct answer is "wine".

In the second question, we will ask you to provide a more detailed answer regarding the item you provided in response to the first question: We will ask you **what kind** of item it was. For example, if your answer to the first question was "wine", you will now be asked "What kind of wine?" The correct answer is "red ".

In the third question, we will ask you to provide a remember/guess judgment regarding the responses that you provided. Choose the option "remember" if the answers you provided to the first and the second questions are based (partially or entirely) on your memory for the slide show. If you have absolutely no memory for these items from the slide show and your responses to the first two questions are mere guesses, choose the option "guess".

Consider your responses carefully. As soon as you press the OK button, you will not be able to change them.

A reminder: You are requested to answer the following questions by relying **solely** on the **slide show** that you viewed in the previous experimental session.

Please do your best and try to avoid answers such as "I don't remember" or "I don't know". If you don't remember a certain item, provide the most suitable answer that you can and give a remember/guess judgment accordingly.

This is the final section of the entire experiment. Please give it your full attention and best efforts. Your responses to the following questions are very important to the goals of this research.

If you wish, you can ask the experimenter to review a printed version of these instructions at any stage. Once you have finished reading these instructions, please raise your hand and call the experimenter.

References

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review, 5*, 1–21.

Belli, R. F., & Loftus, E. F. (1996). The pliability of autobiographical memory: Misinformation and the false memory problem. In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 157–179). New York: Cambridge University Press.

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. C. Solso (Ed.), *Information processing and cognition* (pp. 123–144). New York: Wiley.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35-67). Hillsdale: Erlbaum.

Bornstein, B. H., Liebel, L. M., & Scarberry, N. C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology, 12*, 119–131.

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71*, 81–129.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. In R. Kail & H. W. Reese (Eds.), *Advances in child development and behavior* (pp. 41–100). San Diego: Academic Press.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*, 164–169.

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kingma, J. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development, 55*, 93.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 35*, 1563–1569.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*, 268–276.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*, 633–642.

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition, 29*, 1120–1129.

Chan, J. C. K., & Langley, M. (in press). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*, 553–571.

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science, 20*, 66–73.

Chandler, C. C., & Fisher, R. P. (1996). Retrieval processes and witness memory. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (2nd ed., pp. 493-524).

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.

Christiaansen, R. E., & Ochalek, K. (1983). Editing misleading information from memory: Evidence for the coexistence of original and postevent information. *Memory & Cognition, 11*, 467–475.

Duchastel, P. C., & Nungester, R. J. (1982). Testing effects measured with alternate test forms. *Journal of Educational Research, 75*, 309–313.

Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 29*, 813–825.

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition, 33*, 770–782.

Fisher, R. P. (1995). Interviewing victims and witnesses of crime. *Psychology, Public Policy, and Law, 1*, 732–764.

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield: Charles C. Thomas.

Fisher, R. P., Falkner, K. L., Trevisan, M., & McCauley, M. R. (2000). Adapting the cognitive interview to enhance long-term (35 years) recall of physical activities. *The Journal of Applied Psychology, 85*, 180–189.

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

Goldsmith, M., & Koriat, A. (1999). The strategic regulation of memory reporting: Mechanisms and performance consequences. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 373–400). Cambridge: MIT Press.

Goldsmith, M., Koriat, A., & Weinberg Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131*, 73–95.

Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language, 52*, 505–525.

Goodman, G. S., & Quas, J. A. (2008). Repeated interviews and children's memory: It's more than just how many. *Current Directions in Psychological Science, 17*, 386–390.

Hall, D. F., Loftus, E. F., & Tousignant, J. P. (1984). Postevent information and changes in recollection for a natural event. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 124–141). New York: Cambridge University Press.

Henkel, L. A. (2004). Erroneous memories arising from repeated attempts to remember. *Journal of Memory and Language, 50*, 26–46.

Henry, L. A., & Gudjonsson, G. H. (2004). The effects of memory trace strength on eyewitness recall in children with and without intellectual disabilities. *Journal of Experimental Child Psychology, 89*, 53–71.

Howe, M. L., Courage, M. L., & Bryant-Brown, L. (1993). Reinstating preschoolers memories. *Developmental Psychology, 29*, 854–869.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*, 852–857.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101*, 621–629.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*, 528–558.

Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*, 151–162.

Kuo, T. M., & Hirshman, E. (1996). Investigations of the testing effect. *The American Journal of Psychology, 109*, 451–464.

Lane, S. M., & Zaragoza, M. S. (2007). A little elaboration goes a long way: The role of generation in eyewitness suggestibility. *Memory & Cognition, 35*, 1255–1266.

Lane, S. M., Mather, M., Villa, D., & Morita, S. K. (2001). How events are reviewed matters: Effects of varied focus on eyewitness suggestibility. *Memory & Cognition, 29*, 940–947.

LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology, 67*, 259–266.

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

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361–366.

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.

Marche, T. A. (1999). Memory strength affects reporting of misinformation. *Journal of Experimental Child Psychology, 73*, 45–71.

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*, 54–70.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114*, 1–16.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 11*, 371–385.

McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition, 17*, 423–434.

Memon, A., Zaragoza, M., Clifford, B. R., & Kidd, L. (2010). Inoculation or antidote? The effects of cognitive interview timing on false memory for forcibly fabricated events. *Law and Human Behavior, 34*, 105–117.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533.

Pansky, A. (2010). *Inoculation against forgetting: Advantages of immediate vs. delayed initial testing due to superior verbatim accessibility.* Manuscript submitted for publication.

Pansky, A., & Koriat, A. (2004). The basic-level convergence effect in memory distortions. *Psychological Science, 15*, 52–59.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness recall and testimony. In K. D. Williams & N. Brewer (Eds.), *Psychology and law: An empirical perspective* (pp. 93–150). New York: Guilford.

Paz-Alonso, P. M., & Goodman, G. S. (2008). Trauma and memory: Effects of post-event misinformation, retrieval order, and retention interval. *Memory, 16*, 58–75.

Pezdek, K., & Roe, C. (1995). The effect of memory trace strength on suggestibility. *Journal of Experimental Child Psychology, 60*, 116–128.

Poole, D. A., & White, L. T. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology, 27*, 975–986.

Powell, M. B., Fisher, R. P., Wright, R., Brewer, N., & Williams, K. D. (2005). Investigative interviewing. In *Psychology and law: An empirical perspective* (pp. 11). New York: Guilford.

Rajaram, S., & Barber, S. J. (2008). Retrieval processes in memory. In H. L. Roediger III (Ed.), *Cognitive psychology of memor: Vol. 2. Learning and memory: A comprehensive reference* (pp. 261–283). Oxford: Elsevier.

Reyna, V. F. (1995). Interference effects in memory and reasoning: A fuzzy trace theory analysis. In F. N. Dempster & C. J. Brainerd (Eds.), *New perspectives on interference and inhibition in cognition* (pp. 29–61). New York: Academic Press.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1–75.

Reyna, V. F., & Titcomb, A. L. (1997). Constraints on the suggetibility of eyewitness testimony: A fuzzy-trace theory analysis. In D. G. Payne & F. G. Conrad (Eds.), *Intersections in basic and applied memory research* (pp. 157–174). Mahwah: Erlbaum.

Roediger, H. L., III, McDermott, K. B., & McDaniel, M. A. (in press). Using testing to improve learning and memory. In M. A. Gernsbacher, R. Pew, L. Hough & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 65-74). New York: Worth Publishing.

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.

Roediger, H. L., III, Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies & D. B. Wright (Eds.), *New frontiers in applied memory* (pp. 13–49). Brighton: Psychology Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Saunders, J., & MacLeod, M. D. (2002). New evidence on the suggestibility of memory: The role of retrieval-induced forgetting in misinformation effects. *Journal of Experimental Psychology. Applied, 8*, 127–142.

Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnesia for details about a violent event. *The Journal of Applied Psychology, 73*, 371–377.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30*, 641–656.

Titcomb, A. L., & Reyna, V. F. (1995). Memory interference and misinformation effects. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 263–294). San Diego: Academic Press.

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1–12.

Warren, A. R., & Lane, P. (1995). Effects of timing and type of questioning on eyewitness accuracy and suggestibility. In M. S. Zaragoza, J. R. Graham, et al. (Eds.), *Memory and testimony in the child witness: Applied psychology. Individual, social, and community issues* (Vol. 1, pp. 44–60). Thousand Oaks: Sage.

Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology. Applied, 14*, 50–60.

Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. *The American Psychologist, 55*, 581–598.

Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology. Human Learning and Memory, 6*, 135–144.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*, 571–580.

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.

Zaragoza, M. S., Belli, R. F., & Payment, K. E. (2007). Misinformation effects and the suggestibility of eyewitness memory. In H. Hayne & M. Garry (Eds.), *Do justice and let the sky fall: Elizabeth Loftus and her contributions to science, law, and academic freedom* (pp. 35–63). Mahwah: Erlbaum.