How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics

GERALD ECHTERHOFF and WILLIAM HIRST
New School University, New York, New York

and

WALTER HUSSY
University of Cologne, Cologne, Germany

Previous findings have been equivocal as to whether the postevent misinformation effect on eyewitness memory is reduced by warnings presented *after* the misinformation (postwarnings). In the present research, social postwarnings, which characterize the postevent source as a low-credibility individual, diminished the misinformation effect in both cued recall and recognition tests. Discrediting the source as being either untrustworthy or incompetent was effective (Experiment 1). Also, postwarned participants rated reality characteristics of their memories more accurately than did participants receiving no or high-credibility information about the postevent source (Experiment 2). A social postwarning yielded the same results as an explicit source-monitoring appeal and led to longer response times for postevent items, relative to a no-warning condition (Experiments 3 and 4). The findings suggest that the reduced misinformation effect was due to more thorough monitoring of memory characteristics by postwarned participants, rather than to a stricter response criterion or to enhanced event memory.

Numerous studies have shown that eyewitnesses' memory for an event can be influenced by so-called *postevent misinformation*—that is, erroneous information presented after the occurrence of the witnessed event (e.g., Loftus, 1979; Loftus, Miller, & Burns, 1978; for a review, see Ayers & Reder, 1998). For example, an eyewitness to a traffic accident may falsely remember that a parked car obstructed the driver's view if a cowitness has asserted the existence of that car in a conversation shortly after the accident. As evidence for the misinformation effect has accrued over the years, researchers have started to wonder to what extent people can resist such unwanted biases. To this end, eyewitnesses have

been provided with helpful response options on the memory test (e.g., Lindsay & Johnson, 1989; Zaragoza & Koshmider, 1989; Zaragoza & Lane, 1994) or have been warned, implicitly or explicitly, against misleading influences emanating from a postevent source (e.g., Blank, 1998; Chambers & Zaragoza, 2001; Christiaansen & Ochalek, 1983; Greene, Flynn, & Loftus, 1982; Lindsay, 1990; Underwood & Pezdek, 1998; Wright, 1993). The present research focused on the second of these two options—that is, the use of additional information serving as a warning against misinformation. The main goals were to investigate the effectiveness of warnings given after the misinformation stage and to collect evidence concerning the processes underlying a reduction in the influence of misinformation.

Interest in the mitigation of the misinformation effect is motivated by both practical and theoretical concerns. On the one hand, it is important for legal practitioners to find ways to improve the accuracy of eyewitness reports and to ensure that legal cases are built on reliable evidence (see Wells & Olson, 2003). On the other hand, the successful reduction of the misinformation effect has implications for theoretical accounts of the effect. For instance, if eyewitnesses who are warned of misleading influences stop reporting biased memories, they must be able to access their memories for the original event (e.g., Christiaansen & Ochalek, 1983; Wright, 1993). Thus, it is important to know whether and how the effect of postevent misinformation on eyewitness memory can be decreased or even fully eliminated.

Experiments 1 and 3 were conducted as part of a doctoral dissertation at the Graduate Faculty of Political and Social Science of the New School for Social Research. The dissertation research was supported by a German Academic Exchange Service (DAAD) fellowship to the first author. Experiment 4 was supported by Grant HU 372/5-I from the German Science Foundation (Deutsche Forschungsgemeinschaft) to the third author. Hartmut Blank, Bertram Gawronski, Norbert Groeben, and Birgit Neumann provided valuable comments on earlier drafts. Gratitude is extended to Stephan Groll for his assistance in conducting and analyzing Experiment 2. We are also indebted to Steve Lindsay for kindly providing a copy of the video that was used as the stimulus material in Experiment 2. We thank Simone Dank, Isabel Lindner, René Kopietz, Susan Namini, and Iris Tomiuk for their dedicated help with data collection and administration. Correspondence concerning this article should be addressed to G. Echterhoff, Department of Psychology, University of Bielefeld, P.O. Box/Postfach 100131, D-33501 Bielefeld, Germany (e-mail: gerald.echterhoff@uni-bielefeld.de).

Pre- Versus Postwarnings: Manipulations Before and After the Postevent Stage

It is critical to distinguish whether eyewitnesses are warned *before* or *after* they receive misinformation. The first, prospective type of warning has been referred to as a *prewarning*, whereas the second, retrospective type has been termed *postwarning* in the literature (see Chambers & Zaragoza, 2001). In everyday life, an explicit prewarning can rarely be provided to casual witnesses of a legally investigated event before they may encounter additional postevent information—for instance, another witness's account. Thus, the possibility has to be reckoned with that eyewitnesses have already been exposed to biasing postevent information when they are being interviewed about the witnessed event, leaving only the option of a postwarning.

To date, empirical evidence concerning the effectiveness of postwarnings against the influence of misinformation has been fairly mixed and inconclusive. In some studies, the influence persisted despite postwarnings (Belli, Lindsay, Gales, & McCarthy, 1994; Frost, Ingraham, & Wilson, 2002; Greene et al., 1982; Higham, 1998; Zaragoza & Lane, 1994, Experiment 4), whereas in others it was reduced (Chambers & Zaragoza, 2001; Highhouse & Bottrill, 1995; Meade & Roediger, 2002) or even eliminated by postwarning manipulations (Blank, 1998; Wright, 1993; Zaragoza & Koshmider, 1989). In research reported by Lindsay (1990), a postwarning diminished the misinformation effect only when the event and the postevent source shared many characteristics, not when they shared only a few characteristics.

Unfortunately, but not surprisingly, the existing postwarning studies are comparable only to some extent. Researchers have, for instance, used different ways to present the misinformation (contradicting or supplementing the event information) and different memory tests (recognition, recall, or source memory). Most important, they have employed different types of postwarnings. In most studies, the postwarning took the form of an *explicit* instruction alerting participants more or less directly to discrepancies between the witnessed event and the postevent information. Apparently, postwarnings that are rather unspecific (Belli et al., 1994; Christiaansen & Ochalek, 1983; Frost et al., 2002; Zaragoza & Lane, 1994) or are phrased in the subjunctive mood (Greene et al., 1982; Higham, 1998) are less effective than postwarnings that are specific concerning misleading details (Wright, 1993) or that are spelled out in the indicative mood (e.g., Lindsay, 1990), thus pointing out definite discrepancies between event and postevent information.¹

Furthermore, eyewitnesses are apparently less likely to report misleading items if there is an additional interpersonal or social incentive to resist the influence of misinformation (see Chambers & Zaragoza, 2001). There are at least two avenues to inducing such an additional motivation: First, the misinformation effect has been reduced when the experimenter personally motivated participants to demonstrate that they could not be deceived—that is, to present themselves in a positive light (see Blank, 1998,

p. 492; Zaragoza & Koshmider, 1989, p. 255). Second, eyewitnesses' motivation to resist misinformation may be enhanced when the source providing postevent information is revealed as another individual exerting a biasing influence. As research on reactance has demonstrated, people often try to avoid being controlled by others and to assert internal control over their beliefs and actions, defending their freedom of thought when it appears to be threatened. Hence, if eyewitnesses suspect that another individual may adversely influence their memory, they should be particularly motivated to edit out misinformation on the memory test. We will refer to the latter way of inducing resistance to misinformation as a *social* postwarning.

Although there are some studies that, on the surface, appear to have tested social postwarnings (Chambers & Zaragoza, 2001; Greene et al., 1982; Highhouse & Bottrill, 1995; Meade & Roediger, 2002, Experiment 1), a careful examination indicates that they have simultaneously provided explicit postwarnings as well. Thus, the effect of social and explicit postwarnings could never be assessed independently. For example, in the study by Highhouse and Bottrill, participants not only were told that the postevent source was another, rather untrustworthy participant, but also were explicitly exhorted to report only event information on the memory test. Remarkably, although the explicit part of the postwarning was weak (i.e., unspecific or phrased in the subjunctive mood) in the four extant studies, the rate of misled responses was substantially diminished in three of them (Chambers & Zaragoza, 2001; Highhouse & Bottrill, 1995; Meade & Roediger, 2002, Experiment 1). Thus, it appears that the social part of the postwarning was critical for the reduction of misinformation influence. The present research was designed to determine whether a social postwarning by itself can reduce the influence of misinformation on eyewitness memory.

Explanations for Resistance to the Influence of Misinformation

When eyewitnesses are given a prewarning, they can prepare to receive biasing information and may avoid deep encoding of the postevent information (see Craik & Lockhart, 1972). In fact, misinformation influence has been reduced by explicit prewarnings (e.g., Greene et al., 1982) and social prewarnings (Dodd & Bradshaw, 1980; V. L. Smith & Ellsworth, 1987; Underwood & Pezdek, 1998). Importantly, although an effective prewarning can be attributed to the different encoding of postevent information by warned versus unwarned participants (e.g., shallow vs. deep), an effective postwarning cannot be due to systematic differences in encoding or in storage of postevent items. Clearly, a postwarning can affect only subsequent processes, such as cognitive operations performed during or shortly before the final memory test.

In this article, we consider three possible accounts of processes underlying successful postwarnings. First, a postwarning may induce eyewitnesses to embrace a conservative response strategy and to adopt a stricter response

criterion on a memory test (for such accounts in related areas, see, e.g., Miller & Wolford, 1999). From this perspective, the misinformation effect is diminished because of a lower readiness or tendency to report uncertain items as having been seen previously. Importantly, the adoption of a stricter response threshold should also decrease the rate at which items other than misleading items (such as new or control items) are remembered.

Second, a postwarning may induce eyewitnesses to use techniques to enhance their memory for the witnessed event, allowing them to report the correct event information and to disregard the postevent information. For instance, postwarned eyewitnesses may recreate an internal state similar to the one at the time when they witnessed the target event (see Eich, 1980), or they may use imagery to reinstate the original encoding context, enhancing their access to event details (see Geiselman, 1988; S. M. Smith, 1979). If a reduction of misinformation influence is based on enhanced event memory, the rate of correctly reported event items should be higher relative to a control condition.

Third, eyewitnesses who are postwarned may monitor the source of their memories with increased effort or stringency, as compared with unwarned eyewitnesses (Johnson, Hashtroudi, & Lindsay, 1993; Zaragoza & Lane, 1994). From a source-monitoring perspective, eyewitnesses are influenced by postevent misinformation when they attribute their memories to the wrong source (see Chambers & Zaragoza, 2001; Johnson et al., 1993; Mitchell & Johnson, 2000). Superficial or lax source monitoring during retrieval impedes correct source identification, allowing memories of misleading items to pass as memories of a witnessed event (Johnson et al., 1993; Mitchell & Johnson, 2000). Conversely, when eyewitnesses are explicitly asked to identify the source of their memories, they often report fewer misleading items, relative to a control group (e.g., Lindsay & Johnson, 1989; Multhaup, De Leonardis, & Johnson, 1999; Zaragoza & Lane, 1994, Experiment 3). According to this account, a postwarning reduces the misinformation effect because it induces eyewitnesses to examine the characteristics of candidate memories more closely at the time of test. Thus, eyewitnesses will refrain from reporting items that possess characteristics diagnostic of the postevent source and lack characteristics diagnostic of the original event.

The source attribution view of misinformation resistance has received some empirical support (Chambers & Zaragoza, 2001; Zaragoza & Lane, 1994). Furthermore, the source-monitoring logic has been fruitfully applied to many other areas of memory distortions, such as false memory in the Deese/Roediger-McDermott paradigm (McCabe & Smith, 2002), unconscious plagiarism (e.g., R. L. Marsh, Landau, & Hicks, 1997), confusion of fact and fiction (E. J. Marsh, Meade, & Roediger, 2003), and confabulations (Johnson & Raye, 1998). This body of work suggests that a reduction of misinformation influence is, at least in part, due to more thorough source monitoring.

The Present Research

Four experiments were conducted to investigate whether social postwarnings effectively reduce the postevent misinformation effect on eyewitness memory and to explore underlying mechanisms. All the experiments employed the standard procedure of postevent misinformation studies: The participants first saw the original event (i.e., a video depicting an accident or a burglary), then read a postevent narrative containing nonevent items (misleading items), and finally took a memory test (recognition or cued recall). Nonevent items were plausible (i.e., schema consistent) but did not appear anywhere in the target video (e.g., a bar of soap in the bathroom). These items supplemented rather than contradicted event information (as in Chambers & Zaragoza, 2001; Meade & Roediger, 2002; Zaragoza & Lane, 1994).

Experiment 1 was conducted to investigate whether different kinds of a social postwarning would reduce the eyewitness misinformation effect in a cued recall test. In Experiment 2, the participants were given a recognition test and were asked to rate reality characteristics of their memories. In Experiments 3 (cued recall test) and 4 (recognition test), the effect of a social postwarning was compared with that of an explicit postwarning exhorting the participants to monitor the source of their memories with particular care. In Experiment 4, the response times (RTs) for recognition decisions were compared between the unwarned and the postwarned conditions.

EXPERIMENT 1

Drawing on the literature on source credibility (e.g., Hovland, Janis, & Kelley, 1953), one can distinguish between social warnings that characterize the source as *untrustworthy* and social warnings that characterize the source as *incompetent*. When communicators intentionally render a false version of the original event—that is, when they exhibit a *reporting bias* (Eagly, Wood, & Chaiken, 1978)—they are *untrustworthy*. When communicators lack sufficient knowledge or competence to provide a correct description—that is, when they exhibit a *knowledge bias* (Eagly et al., 1978)—they are *incompetent*. An example of an *untrustworthy* source is a driver who has caused a car accident, who is likely to fabricate a distorted account of the accident (see Dodd & Bradshaw, 1980). An example of an *incompetent* source is an individual who only heard about the accident and provides the postevent information secondhand (see V. L. Smith & Ellsworth, 1987).

Method

Participants. The participants were 91 students at the New School University (63 of them female, 28 male; mean age = 26.6 years). They were paid \$10 for their participation in a study ostensibly about the perception and description of events.

Design. The experiment consisted of a 3×2 mixed design. Type of postwarning (untrustworthy source vs. incompetent source vs. no warning) was manipulated between subjects; item type (misleading vs. control) was manipulated within subjects. The participants in this and all the subsequent experiments were assigned randomly to the postwarning conditions. The primary dependent variable

was the rate of misled responses; other dependent measures were based on recalled event items and on guessed control items.

Materials and Procedure. The participants were tested individually. First, they watched a video in which a woman attends to three young children in her car until she almost hits a boy on a bicycle (1 min 15 sec). After an unrelated 5-min filler task, the participants read a written postevent narrative describing the video (380 words), which had been constructed on the basis of pretests with 30 participants. Each narrative contained four misleading items that were drawn from a pool of eight nonevent items (see the Appendix), which were not shown in the video. The remaining four nonevent items were control items (i.e., items included neither in the video nor in the postevent narrative), which were probed for in the cued recall test to obtain guessing rates. The occurrence of the details as either misleading or control items was counterbalanced across participants. In order to employ different combinations of misleading items, 36 versions of the narrative were constructed.

After an unrelated 10-min filler task, the participants in the two postwarning conditions received bogus information about the source of the narrative. In the *untrustworthy source* condition, they read the following: "You should be aware that the description of the event was Betty's [the driver's] account in court where she had to explain how and why the accident had happened from her point of view." In the *incompetent source* condition, the participants read the following: "You should be aware that the description has been written by a high school student who had to write it as part of an aptitude test and performed below average." The participants in the no-warning condition were asked to think briefly about the narrative but were not given any source information.

Just before the memory test, the postwarned participants were reminded of the alleged source of the postevent narrative. The cued recall memory test contained 16 item probes (four questions probing for misleading items, four questions probing for control items, and eight questions probing for event items). All the probes consisted of a short question (e.g., "In the video, what hung at the handlebar of the boy's bicycle?"). The order of the item probes was randomized, with the constraint that two nonevent item probes should not occur in immediate succession. Finally, a short interview was conducted to confirm that the postwarned participants had understood the social postwarning and accepted the alleged postevent source as a real individual.

Measures. The dependent measure was the rate of actual item responses relative to the total number of possible item responses for each item type. If, for example, a participant reported having seen two misleading items in the video, a rate of $2/4 = .50$ was assigned. An additional score of recall performance, described by Watkins and Gardiner (1982), was calculated for each participant in order to correct for guessing. According to Watkins and Gardiner, "true" recall performance (i.e., recall uncontaminated by guessing) can be estimated using the equation $T = (r - g)/(1 - g)$, where r and g denote the proportions of recalled event items and of guessed control items, respectively. The stimulus material was not counterbalanced between event and control items. However, there was no reason to assume that any effects of item material would differ between the test conditions.

Results and Discussion

In this and all the subsequent experiments, one-tailed statistical tests were employed to determine the effect of a social postwarning; all other tests were two-tailed. The acceptable Type I error probability was $p < .05$.

Misled responses. As is shown in Table 1, the participants given no warning reported misleading items more often ($M = .46$) than did the participants told about an untrustworthy source ($M = .32$) or an incompetent source [$M = .24$; $F(2,88) = 5.00$, $MS_e = 0.08$, $p < .01$]. Im-

portantly, two pairwise planned contrasts revealed that both types of social postwarning led to significantly lower rates of misled responses [$t(88) = 1.97$, $SE = 0.07$, $p < .05$, and $t(88) = 3.10$, $SE = 0.07$, $p < .01$, for untrustworthy and incompetent sources, respectively]. As in many previous studies, the rate of misled responses was compared with the rate of reported control items to determine whether there was an effect attributable to postevent misinformation. A 3 (postwarning: untrustworthy source vs. incompetent source vs. no warning) $\times$ 2 (item type: misleading vs. control) ANOVA revealed a main effect for item type [$F(1,88) = 64.38$, $MS_e = 0.05$, $p < .001$], which was qualified by a significant interaction [$F(2,88) = 7.66$, $p < .001$]. This interaction confirmed that the size of the misinformation effect depended on whether the participants received a social postwarning. No significant main effect was found for type of postwarning [$F(2,88) = 1.73$, n.s.]. The misinformation effect, albeit reduced, was not fully eliminated in the two postwarning conditions: The postwarned participants still reported significantly more misleading items than control items [$F(1,88) = 17.41$, $SE = 0.06$, $p < .001$, and $F(1,88) = 4.18$, $SE = 0.06$, $p < .05$, for the untrustworthy and incompetent source conditions, respectively].

Guessing. Surprisingly, the participants given either type of postwarning appeared to guess more control items ($M = .09$ and $.13$ for untrustworthy and incompetent sources, respectively) than did the participants receiving no warning ($M = .05$). However, these differences failed to be statistically reliable [$F(2,88) = 2.26$, $MS_e = 0.02$, $p = .11$; all $ps > .11$ in Bonferroni post hoc comparisons]. If anything, the response criterion was apparently more lenient in the postwarned conditions. When guessed items other than control items were included (i.e., items *not* from the pool of nonevent items—e.g., *lamp post*), the results of these analyses were confirmed (i.e., the guessing tendency did not differ among the three groups).

Recall performance. There was no evidence that the mean rates of correct responses to event item probes differed depending on the type of postwarning [see Table 1, right columns; $F(2,88) < 1$, n.s.]. Similarly, the adjusted recall score T (Watkins & Gardiner, 1982) did not differ

Table 1
Experiment 1: Mean Recall Rates as a Function of Item Type and Type of Postwarning

Postwarning	Misleading		Control		Event	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No warning	.46	.31	.05	.12	.51	.20
Untrustworthy source	.32	.26	.09	.14	.52	.23
Incompetent source	.24	.26	.13	.15	.51	.18

Note—In the no-warning condition, $n = 34$; in the untrustworthy source condition, $n = 29$; in the incompetent source condition, $n = 28$. Recall rates represent the number of actual item responses in the cued recall test, relative to the total number of possible item responses for each item type. If, for example, a participant reported having seen two of the four misleading items, a score of $2/4 = .50$ was assigned.

reliably between the no-warning ($M = .48$, $SD = .22$), the untrustworthy source ($M = .47$, $SD = .24$), and the incompetent source ($M = .44$, $SD = .19$) conditions [$F(2,88) < 1$, n.s.].

In sum, as compared with a no-warning condition, social postwarnings significantly reduced the misinformation effect in a cued recall test, irrespective of whether they characterized the postevent source as untrustworthy or incompetent. The guessing rate for control items did not differ reliably between the postwarning conditions. The postwarned participants even exhibited a slightly (although not reliably) higher guessing tendency than did the unwarned participants. Also, neither the unadjusted rate of correctly recalled event items nor the guessing-adjusted score for recall performance differed between the postwarned and the unwarned groups. In fact, event recall performance was remarkably similar in all the conditions. These results suggest that the reduction of the misinformation effect was not due to a stricter overall response criterion or to memory-enhancing techniques employed by the postwarned participants.

Can the misinformation effect also be reduced with other test formats—for instance, a recognition test? Although much debate exists over the memory processes tapped by recognition and recall tests, it is safe to assume that the two test formats differ to some extent. For instance, a recognition probe requires a decision about whether an item was seen in the target context, thus implicitly asking for the reconstruction of the source from which the item was acquired (see Tulving, 1976). In accord with this view, some two-process models of recognition memory assume that a first, familiarity-based process is followed by a second, slower process that involves the retrieval of the context or source of test items (see Hintzman, 2000). In contrast, a recall task in a misinformation study does not present participants with misleading items, allowing a self-initiated search for item memories. In this search, postwarned eyewitnesses may not even consider misleading items. Thus, a recognition test is more likely than a recall task to elicit source judgments about misleading items.

To assess source-monitoring efficacy more directly than in the first experiment, we asked the participants in Experiment 2 to rate the characteristics of their memories. Drawing on reality-monitoring cues investigated by Johnson and colleagues (e.g., Johnson, Foley, Suengas, & Raye, 1988), we chose three characteristics that are diagnostic of visually perceived events—that is, amount of visual details, vividness, and clarity of spatial context.

EXPERIMENT 2

Since the two different social postwarnings did not have different effects in the first experiment, in Experiment 2 we employed only one type, which depicted the postevent source as untrustworthy, rather than incompetent. The experiment included an additional condition in which the postevent source was characterized as a highly credible source (i.e., a police officer). Thus, we could de-

termine whether eyewitnesses actually take into account the presumable credibility of the postevent source in the memory test. Also, to increase the likelihood of source judgments in the generation of item responses, a yes–no recognition test was employed.

Method

Participants. The participants were 60 people (50 students) who responded to advertisements placed on the campus of the University of Cologne (40 of them female, 20 male; mean age = 28.5 years). They received curricular credit or a compensation of 8 DM (approximately \$4 at that time).

Design. The experiment consisted of a 3×2 mixed design. Source information (social postwarning vs. social validation vs. no information) was manipulated between subjects; item type (misleading vs. new) was manipulated within subjects. The primary dependent measure was the rate of misled responses. Additional measures were hit and false alarm rates, as well as measures of response bias and discrimination performance.

Materials and Procedure. The materials and procedure were the same as those in Experiment 1, except for the following changes. The video showed a burglar searching a house for valuables after a woman and her daughter had left (approximately 6 min). Before reading the postevent narrative, the participants worked on an unrelated 10-min filler task. The pool of nonevent items was increased from 8 to 16. Eight items were used as misleading items in the postevent narrative, which contained about 900 words. The 8 remaining nonevent items were used as new items in the recognition test. The item pool for this experiment can be obtained at <http://www.uni-koeln.de/psych/methoden/matmisinformation.html>.

The participants in the *social postwarning* condition were told that the postevent source was another participant who had been asked to write an event account in the style of “Baron Münchhausen.” Baron Münchhausen is a well-known figure in German cultural tradition, who is infamous for lying and confabulating. The participants in the *social validation* condition were told that the postevent narrative was written by a police officer who had been asked to write a detailed report about the events in the video. After an unrelated 3-min filler task, the participants were given a yes–no recognition test consisting of 8 misleading items, 8 control items, and 16 event items. The participants were asked to decide whether short phrases (e.g., “There was a pot on the stove”) were true of the video (“yes”) or not (“no”).

If the participants accepted an item, they were also asked to rate the memory characteristics. These ratings were collected using three items from the memory characteristics questionnaire (MCQ; see Johnson et al., 1988). The MCQ is designed to assess differences in qualitative characteristics of memories for real and imagined events. The participants were asked to rate the amount of visual details (“My memory for this item involves visual detail: 1 = little or none, 7 = a lot”), the vividness of their memory (“The vividness of my memory is 1 = not vivid at all, 7 = very vivid”), and the clarity of spatial context information (“Relative spatial arrangement of objects in close proximity to the item is 1 = vague, 7 = clear”).

Two manipulation check ratings, using 9-point scales, were administered to assess the postevent source’s perceived credibility: “Do you think the description was written by a credible person?” (1 = not credible at all, 9 = highly credible) and “Do you think the description of the video correctly rendered the event?” (1 = not at all correct, 9 = very correct).

Measures. Measures of response bias and discrimination performance were calculated on the basis of hit and false alarm rates (see Snodgrass & Corwin, 1988). The discrimination index P_i denotes the probability that an old item will exceed the old recognition threshold or that a new item will exceed the new recognition threshold, respectively ($P_i = H - FA$, where H is the hit rate and FA

is the false alarm rate). The response criterion or bias characterizes a participant's tendency to accept items in a recognition test when he or she is uncertain as to whether the item is old or new [$B_r = FA/(1 - H + FA)$]. A score above .5 indicates a liberal response criterion, whereas a score below .5 represents a strict criterion.

Results and Discussion

Manipulation check. The mean score of the two manipulation check ratings was used as a single measure of the postevent source's perceived credibility. Results showed that the participants receiving the social postwarning rated the postevent source as less credible ($M = 5.22$, $SD = 1.59$) than did the participants receiving the social validation ($M = 6.53$, $SD = 1.77$) or no source information [$M = 6.23$, $SD = 2.33$; $t(57) = 2.23$, $SE = 0.52$, $p < .05$, as calculated with a planned contrast between the postwarning condition and the other two conditions].

False recognition: Misled responses and false alarms. As can be seen in Table 2, the social postwarning reduced the misinformation effect. The participants given no source information ($M = .26$) or a social validation ($M = .31$) falsely recognized misleading items more often than did the participants given a social postwarning [$M = .13$; $t(57) = 2.20$, $SE = 0.07$, $p < .05$, as calculated with a planned contrast between the postwarning condition and the other two conditions]. In contrast, false alarm rates did not differ between these conditions ($F < 1$, n.s.). A 3 (source information: social postwarning vs. social validation vs. no information) $\times$ 2 (item type: misleading vs. new) ANOVA yielded no significant main effect of source information [$F(2,57) = 1.37$, n.s.] and a significant main effect of item type [$F(1,57) = 25.90$, $MS_e = 0.03$, $p < .001$], which was qualified by a significant source information $\times$ item type interaction [$F(2,57) = 3.48$, $p < .05$]. Thus, the size of the misinformation effect differed depending on the type of source information. From a statistical point of view, the misinformation effect was eliminated in the postwarning condition: The participants who were led to regard the postevent source as an unreliable person did not accept more misleading items than new items [$F(1,57) < 1$, n.s., as calculated with a pairwise comparison].

Hits. Hit rates did not differ between the groups of participants given different information about the postevent source (see Table 2, right columns). The participants given the social postwarning did not recognize

more (or fewer) event items ($M = .82$) than did the participants receiving no source information ($M = .82$) or a social validation [$M = .78$; $F(2,57) = 1.08$, n.s.].

Response bias and discrimination performance.

There was no indication that the participants receiving the social postwarning adopted a stricter response criterion B_r ($M = .24$, $SD = .29$) than did the participants given no source information ($M = .19$, $SD = .29$) or the social validation [$M = .18$, $SD = .27$; $F(2,57) < 1$, n.s.]. If anything, the participants given a postwarning displayed an even more liberal response bias, as indicated by the slightly higher score for B_r . Similarly, the participants in the social postwarning condition did not exhibit a better (or worse) discrimination performance P_r ($M = .74$, $SD = .15$) than did the participants given no source information ($M = .76$, $SD = .14$) or the participants led to regard the postevent source as a credible person [$M = .72$, $SD = .11$; $F(2,57) < 1$, n.s.]. Thus, differences in misled responses between the three conditions could not be attributed to differences in the participants' response criterion or underlying recognition performance. Analyses based on signal detection indexes of discrimination (d') and bias (C) led to the same conclusions.

Ratings of memory characteristics. Because the participants rated remembered items (not rejected items), a sufficiently large number of observations was obtained only for correctly recognized event items (not for misleading and new items). First, mean scores for each of the three ratings (visual details, vividness, and spatial context clarity) were calculated for event items within participants. Second, the three mean ratings were averaged to obtain one aggregate rating of reality characteristics for each participant. As is predicted by the source-monitoring account of misinformation reduction, the participants given the social postwarning indicated higher ratings of reality characteristics for their event memories ($M = 4.68$, $SD = 1.28$) than did the participants receiving no source information ($M = 4.01$, $SD = 0.72$) or the social validation [$M = 3.99$, $SD = 0.63$; $t(57) = 2.62$, $SE = 0.25$, $p < .01$, as calculated with a planned contrast between the postwarning and the other two conditions]. [A one-way ANOVA similarly indicated significant differences across the three conditions; $F(2,57) = 3.44$, $MS_e = 0.86$, $p < .05$.] Thus, the participants given the social postwarning exhibited more accurate assessments of the characteristics of their event memories than did the participants in the other conditions.

In sum, Experiment 2 replicated the main findings in Experiment 1 with a recognition test: A social postwarning reduced the misinformation effect on eyewitness memory, relative to the participants receiving no or high-credibility information about the postevent source. The postwarned participants rated the reality characteristics of event items more accurately than did the other participants. This finding suggests that the postwarning induced a more careful scrutiny of source-diagnostic features, which enabled the participants to dismiss misleading items as not having been presented in the target

Table 2
Experiment 2: Mean Recognition Rates as a Function of Item Type and Information About the Postevent Source

Source Information	Misleading		New		Event	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No information	.26	.23	.06	.14	.82	.11
Social postwarning	.13	.19	.08	.11	.82	.11
Social validation	.31	.32	.06	.10	.78	.10

Note—In each condition of source information, $n = 20$. Recognition rates represent the relative frequency of "yes" responses in the recognition test.

event. Whereas higher ratings of reality characteristics in the postwarning condition are, in principle, also consistent with a memory enhancement account, they would not be predicted by a response criterion explanation.

The following two experiments were conducted to obtain more evidence as to whether a reduction of the misinformation influence could be due to stricter source monitoring. To this end, some participants were explicitly exhorted to be aware of possible discrepancies between the event and the postevent account and to attend to the sources of their memories. In Experiment 3, a cued recall test was employed, whereas in Experiment 4, a yes–no recognition test was used to elicit the participants' memories. As in the first two experiments, we wanted to determine whether a reduction of the misinformation effect depends on the format of the memory test.

EXPERIMENT 3

Method

Participants. The participants were 70 people (55 students) who responded to advertisements placed on the campus of the University of Cologne (42 of them female, 28 male; mean age = 29.3 years). They received curricular credit or a compensation of 6 DM (approximately \$3 at that time).

Design. The experiment consisted of a 3×2 mixed design. Postwarning (social postwarning vs. explicit monitoring vs. no warning) was manipulated between subjects; item type (misleading vs. new) was manipulated within subjects. The primary dependent variable was the rate of misled responses; other dependent measures were based on recalled event items and on guessed control items.

Materials and Procedure. The materials and procedure were the same as those in Experiment 1, except for the following modifications. The participants in the *explicit monitoring* condition received the following instruction: "The description of the video you read in the beginning contained some details that were not depicted in the video. Please write down only those items you remember having seen in the video. Do not write down any item that you remember having read only in the description." The participants given a *social postwarning* received the following instruction: "The description of the event is based on what a neighbor of Betty's told a local news reporter. She reported what Betty herself had told her about the incident." (Betty was the driver that caused the accident in the video.) Finally, the same manipulation check was administered as that in Experiment 2.

Results and Discussion

Manipulation check. The postevent source was rated as less credible in the social postwarning condition ($M = 5.15$, $SD = 1.71$) and the explicit monitoring condition ($M = 5.20$, $SD = 1.60$) than in the no-warning condition [$M = 6.15$, $SD = 1.93$; $t(65) = 1.90$, $SE = 0.52$, $p < .05$, and $t(65) = 1.85$, $SE = 0.52$, $p < .05$, respectively, calculated with planned contrasts].

Misled responses. As is shown in Table 3, the participants given no warning reported misleading items more often ($M = .41$) than did the participants given a social postwarning ($M = .24$) or an explicit monitoring instruction [$M = .19$; $F(2,67) = 4.49$, $MS_e = 0.07$, $p < .05$]. Importantly, two pairwise planned contrasts showed that both the social postwarning and the monitoring appeal led to significantly lower rates of misled responses,

relative to the unwarned condition [$t(67) = 2.25$, $SE = 0.08$, $p < .05$, and $t(67) = 2.84$, $SE = 0.08$, $p < .01$, respectively]. The difference between the social warning and the explicit monitoring appeal was not significant [$t(67) = 0.62$, n.s.]. A 3 (postwarning: social postwarning vs. explicit monitoring vs. no warning) $\times$ 2 (item type: misleading vs. control) ANOVA yielded a main effect for item type [$F(1,67) = 43.88$, $MS_e = 0.04$, $p < .001$], which was qualified by a significant interaction [$F(2,67) = 4.07$, $p < .05$]. This interaction confirmed that the size of the misinformation effect differed depending on whether the participants received a warning instruction after the misinformation stage. A significant main effect was also found for postwarning [$F(2,67) = 3.54$, $MS_e = 0.04$, $p < .05$], which was due to the higher rate of recalled (misleading) items in the no-warning condition. As in Experiment 1, the misinformation effect, albeit reduced, was not fully eliminated in the two postwarning conditions: The participants given the social postwarning or the explicit monitoring instruction still reported significantly more misleading items than control items [$F(1,67) = 10.07$, $SE = 0.06$, $p < .01$, and $F(1,67) = 4.92$, $SE = 0.06$, $p < .05$, respectively].

Guessing. The rate of reported control items in the cued recall test differed only by insignificant amounts between the social postwarning condition ($M = .05$), the explicit monitoring condition ($M = .06$), and the no-warning condition [$M = .05$; also see Table 3, middle columns; $F(2,67) < 1$, n.s.; all $ps > .98$ in Bonferroni post hoc comparisons]. As in Experiment 1, the lower rate of misled responses could not be explained by a higher threshold for reporting control items. When guessed items other than the control items were included (i.e., items not from the pool of nonevent items), the results of these analyses were confirmed—that is, the guessing tendency did not differ across the three groups.

Recall performance. No statistically significant differences were obtained for the mean rates of correct responses to event item probes [see Table 3, right columns; $F(2,67) = 1.01$, n.s.]. Similarly, the guessing-adjusted recall score T (Watkins & Gardiner, 1982; see Experiment 1 above) did not differ significantly between the no-warning ($M = .60$, $SD = .23$), the untrustworthy source ($M = .53$, $SD = .20$), and the incompetent source ($M = .53$, $SD = .22$) conditions [$F(2,67) < 1$, n.s.]. Thus, the lower rate of misled responses in the social

Table 3
Experiment 3: Mean Recall Rates as a Function of Item Type and Type of Postwarning

	Misleading		Control		Event	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Postwarning						
No warning	.41	.26	.05	.11	.65	.19
Social postwarning	.24	.27	.05	.10	.58	.17
Explicit monitoring	.19	.27	.06	.12	.62	.16

Note—In the no-warning condition, $n = 23$; in the social postwarning condition, $n = 24$; in the explicit monitoring condition, $n = 23$. See the note to Table 1 and the text for further explanations.

postwarning condition could not be due to enhanced recall of the target event. If anything, recall performance was even worse for the participants given the social postwarning or asked to monitor the source of their memories more carefully.

In sum, when the participants were postwarned, they reported fewer misleading items on the final cued recall test than did the unwarned participants. As in the previous two experiments, additional analyses demonstrated that the reduction of the misinformation influence could not be attributed to a stricter overall response criterion or to enhanced memory for the witnessed event. Furthermore, the participants who were expressly asked to perform more careful source monitoring provided the same pattern of responses as did the participants given a social postwarning. Arguably, the explicit monitoring appeal induced people to attend more closely to the sources of their memories. Although these findings do not directly reveal source scrutiny in the social postwarning group, they do indicate that the reduction of the misinformation effect is, at least to some extent, based on a more careful examination of source-diagnostic characteristics.

In accordance with our rationale for the second experiment, Experiment 4 replicated the previous experiment with a yes–no recognition test, to increase the likelihood of judgments about the source of the items (see Hintzman, 2000). Also, RTs were measured for each recognition response. In many misinformation studies, misleading items are embedded in a postevent narrative that is detailed and mostly accurate. Thus, assuming that an item was provided by the postevent source does not resolve the question of whether the item was also witnessed in the original event (see Chambers & Zaragoza, 2001). Determining whether one can report a postevent item should require a more extensive and thorough evaluation of source-diagnostic characteristics of the candidate memory. Only such an additional scrutiny of memory characteristics (such as visual details, vividness, or clarity of spatial context) will confirm or disconfirm whether the item was seen in the target event. Hence, a postwarning should lead to longer RTs for postevent items, relative to an unwarned condition.

EXPERIMENT 4

Method

Participants. The participants were 88 students at the University of Cologne (61 of them female, 27 male; mean age = 25.6 years). For their participation, they received either curricular credit or a compensation of €8 (approximately \$8 at the time).

Design. The experiment consisted of a 3×2 mixed design. Postwarning (social postwarning vs. explicit monitoring vs. no warning) was manipulated between subjects; item type (misleading vs. new) was manipulated within subjects. In addition to the dependent measures used in Experiment 2, we also calculated mean RTs.

Materials and Procedure. The materials and procedure were the same as those in Experiment 3, except for the following modifications. The video (about 12 min in length) showed a burglar searching a house for valuables after a resident (a young woman) had left. The participants completed a 2-min filler task after watching the video and then read the postevent narrative (approximately

1,100 words). On the basis of pretests, a pool of 8 nonevent items was constructed (available at <http://www.uni-koeln.de/psych/methoden/matmisinformation.html>). After another 5-min filler task, the participants received the same social postwarning as that in Experiment 2 and took a yes–no recognition test. The test was run on PCs (connected to 15-in. color monitors) using the software MediaLab/Direct RT. After the presentation of two warm-up items, which were not included in the analyses, 40 test items (4 misleading, 4 new, 16 event-only, and 16 event-plus-postevent items) were presented in a fully randomized order. Each item remained on the screen until the participant responded. RTs were recorded for all the test items. Finally, the same manipulation check as that in Experiments 2 and 3 was administered, this time on a 7-point scale (instead of a 9-point scale).

Measures. Response bias and discrimination performance were calculated as in Experiment 2. RTs were analyzed for both “yes” and “no” responses. Outliers—that is, RTs longer than a conventional threshold (mean + 2 * standard deviation, calculated separately for each item)—were eliminated from the analyses. By this procedure, 3.3% of all the RTs were excluded (the outlier percentage was 2.9% in the no-warning condition, 3.5% in the postwarning condition, and 3.4% in the explicit monitoring condition).

Results and Discussion

Manipulation check. Ratings of the postevent source’s credibility were lower in the social postwarning condition ($M = 3.58$, $SD = 1.22$) and in the explicit monitoring condition ($M = 3.52$, $SD = 1.25$) than in the no-warning condition [$M = 4.23$, $SD = 1.41$; $t(85) = 1.90$, $SE = 0.34$, $p < .05$, and $t(85) = 2.07$, $SE = 0.34$, $p < .05$, respectively].

False recognition: Misled responses and false alarms. Again, the misinformation effect was effectively reduced (see Table 4). The participants given a social postwarning ($M = .26$) or an explicit monitoring appeal ($M = .23$) falsely recognized misleading items less often than did the participants given no warning [$M = .42$; $F(2,85) = 4.30$, $MS_e = 0.07$, $p < .05$]. Two pairwise planned contrasts revealed that both the social postwarning and the monitoring instructions led to significantly lower rates of misled responses, relative to the unwarned condition [$t(85) = 2.27$, $SE = 0.07$, $p < .05$, and $t(85) = 2.75$, $SE = 0.07$, $p < .01$, respectively]. As in Experiment 3, the difference between the social postwarning and the explicit monitoring conditions was small and statistically nonsignificant [$t(85) = 0.50$, n.s.]. False alarm rates did not differ between the three conditions ($F < 1$, n.s.).

Table 4
Experiment 4: Mean Recognition Rates as a Function of Item Type and Type of Postwarning

Postwarning	Misleading		New		Event	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No warning	.42	.30	.22	.27	.54	.16
Social postwarning	.26	.27	.22	.23	.49	.12
Explicit monitoring	.23	.20	.19	.21	.47	.15

Note—In the no-warning condition, $n = 29$; in the social postwarning condition, $n = 30$; in the explicit monitoring condition, $n = 29$. Recognition rates represent the relative frequency of “yes” responses in the recognition test.

A 3 (postwarning: social postwarning vs. explicit monitoring vs. no warning) $\times$ 2 (item type: misleading vs. new) ANOVA yielded a significant main effect of item type [$F(1,85) = 10.48$, $MS_e = 0.04$, $p < .01$] and a marginally significant postwarning $\times$ item type interaction [$F(2,85) = 2.83$, $p < .07$]. Apparently, the size of the misinformation effect was attenuated when the participants received a social postwarning or an explicit monitoring instruction. Postwarning did not have a significant effect [$F(2,85) = 2.26$, $MS_e = 0.09$, $p = .11$]. From a statistical point of view, the misinformation effect was eliminated in both the social postwarning and the explicit monitoring conditions: The participants in each of these conditions did not accept more misleading items than new items (both $F_s < 1$, n.s., as calculated with pairwise comparisons).

Hits. Hit rates (see Table 4, right columns) did not differ reliably across the three postwarning conditions [$F(2,85) = 2.11$, $MS_e = 0.02$, $p = .13$]. Thus, the lower rate of misled responses in the postwarned conditions was apparently not accompanied by a higher rate of correct event item recognition. If anything, the unwarned participants recognized even more event items ($M = .54$) than did the participants receiving no postwarning ($M = .49$) or an explicit monitoring appeal ($M = .47$).

Response bias and discrimination performance. Differences in response bias (B_r) between the social postwarning condition ($M = .25$, $SD = .24$), the explicit monitoring condition ($M = .24$, $SD = .25$), and the no-warning condition ($M = .26$, $SD = .29$) were small and statistically not reliable ($F < 1$, n.s.). Similarly, discrimination performance (P_r) was not significantly better or worse when the participants were given the social postwarning ($M = .27$, $SD = .22$), the explicit monitoring instruction ($M = .28$, $SD = .16$), or no warning ($M = .32$, $SD = .25$; $F < 1$, n.s.). Thus, differences in misled responses between the three conditions could not be attributed to differences in the participants' response criterion or underlying discrimination performance. (As in Experiment 2, analyses based on signal detection indexes of discrimination [d'] and bias [C] led to the same conclusions.) The relatively low discrimination score in the whole sample could be due to the item material, which was apparently more difficult to remember than the material in the previous experiments. This increased difficulty may have been a result of the more complex and longer video that was presented as the target event.

Response times. Mean RTs for postevent items (misleading items and event items that were also mentioned in the postevent narrative) were calculated for each participant, including both correct and incorrect responses. These mean RTs were adjusted by the individual baseline RT for all other test items (i.e., new and event-only items) to control for individual differences (such as participants' different reading or motor skills) and for irrelevant differences possibly induced by the postwarning manipulation (such as more careful or slower reading). The adjustment was based on a regression with the mean

RTs for postevent items as the dependent measure and the individual baseline RTs as the independent measure ($\beta = .51$), yielding residual scores devoid of the irrelevant variance. For ease of interpretation, a constant (i.e., the mean RT for postevent items across the three postwarning conditions) was added to the residual scores. Importantly, the participants given the social postwarning ($M = 4,364$ msec, $SD = 1,054$) or the explicit monitoring condition ($M = 4,524$ msec, $SD = 1,194$) exhibited longer RTs for postevent items than did the participants receiving no warning [$M = 3,752$ msec, $SD = 1,017$; $F(2,85) = 4.05$, $MS_e = 1,189,958$, $p < .05$]. Pairwise contrasts confirmed that RTs in either the social postwarning or the explicit monitoring condition were longer than those in the no-warning condition [$t(85) = 2.15$, $SE = 284.1$, $p < .05$, and $t(85) = 2.69$, $SE = 286.5$, $p < .01$, respectively]. When the analyses were restricted to "yes" responses or when outliers were recoded to an average outlier threshold (i.e., 10,000 msec), the analyses led to the same conclusion.

In sum, Experiment 4 replicated the findings of Experiment 3 with a recognition test. The participants given a social postwarning resisted the influence of misinformation to the same extent as the participants exhorted to perform stricter source monitoring, relative to an unwarned control group. Also, the postwarned participants took longer than did the unwarned participants to respond to items mentioned in the postevent narrative. If we assume that an additional scrutiny of memory characteristics is needed to determine whether a postevent item was seen in the target event or not (Chambers & Zaragoza, 2001), the RT findings suggest that postwarnings instigate stricter monitoring of memory characteristics.

GENERAL DISCUSSION

Whereas previous research had yielded equivocal results as to whether the postevent misinformation effect on eyewitness memory can be reduced by retrospective warnings, the present series of experiments sends out an unambiguous message: Postwarnings against the influence of postevent misinformation were clearly effective. In all four experiments, the misinformation effect on eyewitness memory was reduced by means of different types of postwarning. Thus, to lower the likelihood and even to prevent eyewitnesses from remembering suggested or false details from a postevent account, they do not have to be warned before or during acquisition of postevent information.

The following findings were relevant to an explanation of how eyewitnesses manage to resist misinformation influence. First, in none of the four experiments was a lower rate of misled responses accompanied by a stricter criterion to report control items or by improved memory for the target event. Second, as compared with a control condition, postwarned eyewitnesses exhibited a more accurate sense of memory characteristics, assigning higher ratings of visual details, vividness, and clarity of spatial

context to event items (Experiment 2). Third, RTs for postevent items were longer in a postwarned than in an unwarned group, suggesting that postwarned eyewitnesses performed an additional scrutiny of memory characteristics to decide whether they had seen these items in the target event (Experiment 4). Fourth, when the format of the memory test (yes–no recognition) required the reconstruction of the previous acquisition context (i.e., the source) and was, hence, more likely to elicit source judgments than were other test formats (cued recall), the misinformation effect was fully eliminated (Experiments 2 and 4).

No aspect of our results suggests that postwarned eyewitnesses adopt a stricter overall response threshold or use techniques (such as generating helpful retrieval cues) to enhance their memory for a witnessed event. It appears that the sum of the findings can be reconciled best with a source-monitoring account of misinformation resistance (Chambers & Zaragoza, 2001; Johnson et al., 1993). According to this approach, eyewitnesses edit out misinformation to the extent that they examine memory characteristics that are diagnostic of the possible sources more thoroughly. Apparently, postwarnings can be one effective way to motivate eyewitnesses to devote more effort to source monitoring. This view is consistent with findings showing that the quality of source judgments at retrieval improves when attentional resources are increased (e.g., Dodson, Holland, & Shimamura, 1998), when working memory capacity is high, rather than low (McCabe & Smith, 2002), or when decision criteria at test are stringent, rather than lax (e.g., Dodson & Johnson, 1993).

In the present research, response bias and memory performance were applied to item memory concerning a single source (i.e., the witnessed event). Thus, the response bias and discrimination indexes referred to decisions about whether an item was old (i.e., seen vs. not seen in the target event). Beyond this conventional application, both bias and discrimination parameters can also be applied to source memory, yielding the probability that an item from a source is correctly remembered versus guessed as being from that source (see Bayen, Murnane, & Erdfelder, 1996). Source discrimination and source bias parameters can be estimated independently from old/new bias and old/new discrimination in multinomial models of source monitoring (Bayen et al., 1996). The use of such models in future research may allow a more refined assessment of how response biases and discrimination performance contribute to eyewitnesses' resistance to misinformation.

Given that the present research invoked source monitoring as an account of misinformation resistance, readers may wonder why our experiments did not employ direct tests of source memory (such as those in Chambers & Zaragoza, 2001; Lindsay & Johnson, 1989; Meade & Roediger, 2002). Indeed, if postwarned eyewitnesses perform more careful source monitoring, they should exhibit improved source memory. However, there is one

important caveat to the practicality of employing an explicit source memory test in misinformation studies: Previous research has shown that when participants are asked to specify the sources of test items, the misinformation effect is substantially reduced relative to conventional test formats, such as recognition (Lindsay & Johnson, 1989; Multhaup et al., 1999; Zaragoza & Lane, 1994, Experiment 3). Hence, the use of an explicit source memory test may preclude the detection of any postwarning effects. In fact, in an unpublished study replicating Experiment 4, we did find that the misinformation effect was reduced by a source memory test (relative to a condition employing a recognition test).

Relation to Other Discounting Paradigms

How do our experiments relate to other paradigms that have been used to investigate whether people can ignore unwanted information? We will briefly discuss two areas of research. First, investigations of information discounting have demonstrated that biased beliefs, once implanted in people's minds, persevere despite debiasing instructions (e.g., Ross, Lepper, & Hubbard, 1975) and that people have a hard time disregarding inadmissible evidence (for a review, see Golding & Long, 1998). However, research on information discounting differs from the present paradigm to the extent that it has focused on judgments (such as guilty verdicts) as a dependent measure. Judgment tasks require people to integrate information into a coherent whole and to draw inferences from the input material. In contrast, memory tasks in misinformation studies involve the retrieval and monitoring of single pieces of input information. Some evidence suggests that the ability to disregard unwanted information in one's judgment may be independent of one's memory for the original information (e.g., Wyer & Unverzagt, 1985). Thus, whenever a test requires people to report a highly integrated representation rather than single pieces of information, people may not rely on careful source monitoring to discount unwanted influences but on other discounting strategies, such as correction by contrast (see, e.g., Wilson & Brekke, 1994).

Second, research on directed forgetting has shown that people are able to suppress unwanted information (see MacLeod, 1998). If one regards the postevent narrative in our experiments as analogous to a to-be-forgotten item list, our postwarnings can be construed as forget instructions in directed forgetting studies using the list method. However, whereas in misinformation studies there is considerable overlap between the items from the two sources, there is no overlap between the to-be-remembered and the to-be-forgotten lists in the typical directed forgetting study. Because of this overlap, it should be more difficult to discriminate the items from different sources in misinformation than in directed forgetting studies. After all, it is not appropriate for eyewitnesses to exclude all items acquired from the postevent source in the recognition test, whereas it is perfectly appropriate for participants in directed forgetting studies to exclude

all items from the F list (MacLeod, 1998). Hence, source-monitoring processes may be more critical in eyewitness postwarnings than in directed forgetting.

Effect and Validity of Social Postwarnings

The present research has provided new evidence concerning a socially framed postwarning, which previously had not been investigated in its own right. Clearly, the participants in our experiments profited from a social postwarning. The magnitude of misinformation influence apparently depended on the perceived credibility of the postevent source: All manipulation checks revealed that the participants given a social postwarning perceived the postevent source as less credible than did the unwarned participants. Also, Experiment 2 showed that the participants regarding the postevent source as a highly credible individual exhibited the standard misinformation effect.

As to the ecological validity of the present research, one may argue that outside the laboratory, credibility cues concerning the postevent source, such as suspicious nonverbal behavior, are typically available during or before the receipt of postevent information. Recipients of communication often pay attention to credibility cues (see, e.g., Zuckerman, DePaulo, & Rosenthal, 1981), which may indeed serve as effective prewarnings against misinformation influence (see Dodd & Bradshaw, 1980; V. L. Smith & Ellsworth, 1987; Underwood & Pezdek, 1998). However, credibility cues can be ambiguous (Johnson, Bush, & Mitchell, 1998). Also, eyewitnesses may not question the reliability of postevent information until they receive additional credibility information. For instance, a cowitness may appear as a fairly credible source until it is disclosed that he or she was not wearing his or her glasses at the time of the witnessed event (Hatvany & Strack, 1980). Thus, experimental studies of social postwarnings do have potential validity outside the confines of the experimental setting. Future research should investigate the effects of social postwarnings with actual individuals as postevent sources, moving from the implied social presence of the source into the domain of face-to-face interaction (see Gabbert, Memon, Allan, & Wright, 2004; Meade & Roediger, 2002).

The implications of this research should be of particular interest to the recently growing community of scholars exploring the social conditions of memory distortion (e.g., Bless & Strack, 1998; Echterhoff & Hirst, 2002; Hirst & Manier, 1996; Meade & Roediger, 2002). Future studies should investigate the specific processes of source monitoring induced by social and explicit postwarnings. Possibly, eyewitnesses given different types of postwarnings use somewhat different memory features or assign different weights to feature dimensions (see Echterhoff & Hussy, 2004; R. L. Marsh & Hicks, 1998). Research along these lines could both enhance our knowledge of source monitoring and allow a detailed investigation of the cognitive processes mediating social influences on memory.

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.
- BAYEN, U. J., MURNANE, K., & ERDFELDER, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 197-215.
- BELLI, R. F., LINDSAY, D. S., GALES, M. S., & MCCARTHY, T. T. (1994). Memory impairment and source misattribution in postevent misinformation experiments with short retention intervals. *Memory & Cognition*, *22*, 40-54.
- BLANK, H. (1998). Memory states and memory tasks: An integrative framework for eyewitness memory and suggestibility. *Memory*, *6*, 481-529.
- BLESS, H., & STRACK, F. (1998). Social influence on memory. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 90-106). London: Sage.
- 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.
- 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.
- CRAIK, F. I. M., & LOCKHART, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671-684.
- DODD, D. H., & BRADSHAW, J. T. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning & Verbal Behavior*, *19*, 695-704.
- DODSON, C. S., HOLLAND, P. W., & SHIMAMURA, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 1121-1136.
- DODSON, C. S., & JOHNSON, M. K. (1993). The rate of false source attributions depends on how questions are asked. *American Journal of Psychology*, *106*, 541-557.
- EAGLY, A. H., WOOD, W., & CHAIKEN, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality & Social Psychology*, *36*, 424-435.
- ECHTERHOFF, G., & HIRST, W. (2002). Remembering in a social context: A conversational view of the study of memory. In G. Echterhoff & M. Saar (Eds.), *Kontexte und Kulturen des Erinnerns: Maurice Halbwachs und das Paradigma des kollektiven Gedächtnisses* [Contexts and cultures of remembering: Maurice Halbwachs and the paradigm of collective memory] (pp. 75-101). Konstanz, Germany: Universitätsverlag Konstanz.
- ECHTERHOFF, G., & HUSSY, W. (2004). Strategies of source attribution: Semantic features and trace strength as cues to the origin of memories. *Swiss Journal of Psychology*, *63*, 93-106.
- EICH, J. E. (1980). The cue-dependent nature of state-dependent retrieval. *Memory & Cognition*, *8*, 157-173.
- 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.
- GABBERT, F., MEMON, A., ALLAN, K., & WRIGHT, D. B. (2004). Say it to my face: Examining the effects of socially encountered misinformation. *Legal & Criminological Psychology*, *9*, 215-227.
- GEISELMAN, R. E. (1988). Improving eyewitness memory through mental reinstatement of context. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 245-266). Chichester, U.K.: Wiley.
- GOLDING, J. M., & LONG, D. L. (1998). There's more to intentional forgetting than directed forgetting: An integrative review. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 59-102). Mahwah, NJ: Erlbaum.
- GREENE, E., FLYNN, M. S., & LOFTUS, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning & Verbal Behavior*, *21*, 207-219.

- HATVANY, N., & STRACK, F. (1980). The impact of a discredited key witness. *Journal of Applied Social Psychology*, **10**, 490-509.
- HIGHAM, P. A. (1998). Believing details known to have been suggested. *British Journal of Psychology*, **89**, 265-283.
- HIGHHOUSE, S., & BOTTRILL, K. V. (1995). The influence of social (mis)information on memory for behavior in an employment interview. *Organizational Behavior & Human Decision Processes*, **62**, 220-229.
- HINTZMAN, D. L. (2000). Memory judgments. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 165-177). Oxford: Oxford University Press.
- HIRST, W., & MANIER, D. (1996). Remembering as communication: A family recounts its past. In D. C. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 271-290). Cambridge: Cambridge University Press.
- HOVLAND, C. I., JANIS, I. L., & KELLEY, M. M. (1953). *Communication and persuasion*. New Haven, CT: Yale University Press.
- JOHNSON, M. K., BUSH, J. G., & MITCHELL, K. J. (1998). Interpersonal reality monitoring: Judging the sources of other people's memories. *Social Cognition*, **16**, 199-224.
- JOHNSON, M. K., FOLEY, M. A., SUENGAS, A. G., & RAYE, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, **117**, 371-376.
- JOHNSON, M. K., HASHTRUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.
- JOHNSON, M. K., & RAYE, C. L. (1998). False memories and confabulation. *Trends in Cognitive Sciences*, **2**, 137-145.
- LINDSAY, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1077-1083.
- LINDSAY, D. S., & JOHNSON, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, **17**, 349-358.
- LOFTUS, E. F. (1979). The malleability of human memory. *American Scientist*, **67**, 312-320.
- LOFTUS, E. F., MILLER, D. G., & BURNS, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 19-31.
- MACLEOD, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1-57). Mahwah, NJ: Erlbaum.
- MARSH, E. J., MEADE, M. L., & ROEDIGER, H. L., III (2003). Learning facts from fiction. *Journal of Memory & Language*, **49**, 519-536.
- MARSH, R. L., & HICKS, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1137-1151.
- MARSH, R. L., LANDAU, J. D., & HICKS, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 886-897.
- MCCABE, D. P., & SMITH, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, **30**, 1065-1077.
- MEADE, M. L., & ROEDIGER, H. L., III (2002). Explorations in the social contagion of memory. *Memory & Cognition*, **30**, 995-1009.
- MILLER, M. B., & WOLFORD, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review*, **106**, 398-405.
- MITCHELL, K. J., & JOHNSON, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179-195). Oxford: Oxford University Press.
- MULTHAUP, K. S., DE LEONARDIS, D. M., & JOHNSON, M. K. (1999). Source memory and eyewitness suggestibility in older adults. *Journal of General Psychology*, **126**, 74-84.
- ROSS, L., LEPPER, M. R., & HUBBARD, M. (1975). Perseverance in self-perception and social perception: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality & Social Psychology*, **35**, 817-829.
- SMITH, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 68-86.
- SMITH, V. L., & ELLSWORTH, P. C. (1987). The social psychology of eyewitness accuracy: Leading questions and communicator expertise. *Journal of Applied Psychology*, **72**, 292-300.
- SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.
- STAPEL, D. A., MARTIN, L. L., & SCHWARZ, N. (1998). The smell of bias: What instigates correction processes in social judgments? *Personality & Social Psychology Bulletin*, **24**, 797-806.
- TULVING, E. (1976). Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 37-73). London: Wiley.
- UNDERWOOD, J., & PEZDEK, K. (1998). Memory suggestibility as an example of the sleeper effect. *Psychonomic Bulletin & Review*, **5**, 449-453.
- WATKINS, M. J., & GARDINER, J. M. (1982). Cued recall. In C. R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 173-195). New York: Academic Press.
- WELLS, G. L., & OLSON, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, **54**, 277-295.
- WILSON, T. D., & BREKKE, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, **116**, 117-142.
- WRIGHT, D. B. (1993). Misinformation and warnings in eyewitness testimony: A new testing procedure to differentiate explanations. *Memory*, **1**, 153-166.
- WYER, R. S., & UNVERZAGT, W. H. (1985). Effects of instructions to disregard information on its subsequent recall and use in making judgments. *Journal of Personality & Social Psychology*, **48**, 533-549.
- ZARAGOZA, M. S., & KOSHMIDER, J. W., III (1989). Misled subjects may know more than their performance implies. *Journal of Experimental Psychology: Learning, Memory, & 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, & Cognition*, **20**, 934-945.
- ZUCKERMAN, M., DEPAULO, B. M., & ROSENTHAL, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1-59). New York: Academic Press.

NOTE

1. In the domain of judgmental debiasing, Stapel, Martin, and Schwarz (1998) have proposed a distinction similar to the one between a subjunctive and an indicative postwarning. According to Stapel et al., an *unconditional warning* unmistakably states the presence of a biasing influence and exhorts participants to correct for this bias under all circumstances. In contrast, a *conditional warning* asks participants to correct for a bias if they feel it is necessary. The authors found that a conditional postwarning led to effective debiasing only when the bias source was salient, whereas an unconditional postwarning was effective irrespective of the salience of the bias source. Similarly, as was indicated by a review of eyewitness misinformation studies, explicit postwarnings phrased in the indicative mood ("There were some crucial differences between the video and the narrative . . .") have more effectively reduced misinformation influence than have postwarnings phrased in the subjunctive mood ("There may have been some crucial differences . . .").

(Continued on next page)

APPENDIX
Pool of Eight Nonevent Items Used as Misleading Items in the Postevent Description,
Included With the Passages They Appeared In (Experiments 1 and 3)

Lucy was sitting in the front seat next to Betty. A *shopping bag* was placed before Lucy on the floor by her feet.

It was a very cold day, so all four of them were wearing woolen hats. Lucy was even wearing *gloves*.

Betty noticed a speed limit sign and started looking at her speedometer that was next to the *fuel gauge*.

Tom and Mary were rather unruly that day and were playing around with a *teddy bear* in the back seat.

Betty started paying more attention to the road and noticed a *flyer* from the shopping mall behind the windshield wiper.

. . . Betty was driving down the snow-covered street. The car went over a *patch of leaves*.

At first, Betty did not realize that a boy in a blue jacket was riding his bike down a driveway and behind a *tree* in the front yard.

. . . she couldn't swerve out of the way because there was a *blue car* coming toward them in the opposite lane.

Note—The passages suggested the existence of an object or detail that was not shown in the original event (i.e., the video). The respective nonevent items are italicized. Materials used in Experiments 2 and 4 are available online at <http://www.uni-koeln.de/psych/methoden/matmisinformation.html>.

(Manuscript received December 2, 2002;
revision accepted for publication August 26, 2004.)