Originally published in: Journal of Applied Research in Memory and Cognition 3 (2014) 77–88
http://dx.doi.org/10.1016/j.jarmac.2014.03.005
This current archived file includes all essential corrections made to the journal article proofs.

How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies

Hartmut Blank[1] **Céline Launay**[2]

[1] University of Portsmouth

[2] University of Toulouse

Corresponding author: Dr. Hartmut Blank, Department of Psychology, University of Portsmouth, King Henry Building, King Henry I Street, Portsmouth PO1 2DY, United Kingdom; E-mail: hartmut.blank@port.ac.uk

Abstract

Four decades of research and hundreds of studies speak to the power of post-event misinformation to bias eyewitness accounts of events (see e.g. Loftus' summary, 2005). A subset of this research has explored if the adverse influence of misinformation on remembering can be undone or at least reduced through a later warning about its presence. We meta-analyzed 25 such post-warning studies (including 155 effect sizes) to determine the effectiveness of different types of warnings and to explore moderator effects. Key findings were that (1) post-warnings are surprisingly effective, reducing the misinformation effect to less than half of its size on average. (2) Some types of post-warning (following a theoretical classification) seem to be more effective than others, particularly studies using an *enlightenment* procedure (Blank, 1998). (3) The post-warning reduction in the misinformation effect reflects a specific increase in misled performance (relative to no warning), at negligible cost for control performance. We conclude with a discussion of theoretical and practical implications.

[77]

[78]

1. Introduction

Pioneering research by Elizabeth Loftus and colleagues has exposed the vulnerability of eyewitness reports to the biasing influence of post-event misinformation (while eyewitness suggestibility more generally has been noted earlier; see Sporer, 1982, for a historical overview). In a prototypical study (e.g. Loftus, Miller, & Burns, 1978), participants are first shown a video or slide sequence of a staged realistic event of some forensic relevance (e.g., a traffic accident or a crime) and are later exposed to misinformation about this event. This can be achieved through 'hiding' misinformation in

apparently neutral questions (e.g., "Did another car pass the red Datsun while it was stopped at the stop sign" – the presupposition here is that there was a stop sign at the intersection, rather than the original yield sign in the slide sequence) or through embedding it in an apparently trustworthy narrative account of the event. Finally, the participants undergo a memory test designed to probe their memory for original event details and/or their endorsement of misleading details. Different memory tests focus on one of two major possible manifestations of misinformation influence (cf. Higham, 1998; Pansky, Tenenboim, & Bar, 2011): (1) poorer memory performance for original event details (e.g. yield sign) that have been the target of post-event misinformation (e.g., stop sign), relative to a no-misinformation control condition; this has been demonstrated using forced-choice recognition (e.g., Loftus et al., 1978), yes-no recognition (e.g., Belli, 1989) or cued recall tests (e.g. Geiselman, Fisher, Cohen, Holland, & Surtes, 1986). Alternatively, or sometimes in addition, (2) researchers have demonstrated stronger endorsement or incorporation of suggested misleading details in memory tests, typically in cued recall and in yes-no recognition but also in source monitoring tests, where participants often mistakenly claim to have encountered a suggested detail in the original event (e.g. Higham, 1998; Lindsay, 1990; Zaragoza & Lane, 1994).[1]

The overwhelming majority of literally hundreds of studies of the eyewitness misinformation effect confirm its existence (in one or both of the forms described above; see Belli & Loftus, 1996; Loftus, 2005; Zaragoza, Belli & Payment, 2006; for overviews). The magnitude of the effect in a given study depends of course on study characteristics and on the nature of the memory test, but even with the most 'conservative' test (McCloskey & Zaragoza's, 1985, modified test procedure) a small but reliable misinformation effect has been found (see Payne, Toglia & Anastasi's meta-analysis, 1994).

Still, this does not mean that the misinformation effect must be accepted as some sort of curse thrust upon memory. Soon after its initial demonstration, researchers have started to look for conditions under which the misinformation is weakened or does not materialize at all. One of the earliest demonstrations along these lines was a study by Dodd and Bradshaw (1980) in which the effect basically disappeared when the misinformation was presented as coming from a biased source (the lawyer representing the driver in a car accident). Following a similar rationale, other researchers employed different forms of *warnings* in order to discourage participants from relying too much on the post-event information and the misleading details contained in it. To our knowledge, Greene, Flynn, and Loftus (1982) were the first to explore the moderating impact of a (mild) warning on the misinformation effect. In their study, some participants were told that "the police cadet who wrote the report [i.e., the post-event narrative containing misinformation; our addition] was inexperienced"; this happened either before or after the presentation of this report. Greene et al. report that only the pre-warning but not the post-warning reduced the misinformation effect (but even so it did not fully eliminate it).

Other researchers explored different types of warnings, partly as required by the specific purposes of their studies. For example, Wright (1993) used an extreme form of warning in which the misleading detail was explicitly *named* and it was made clear that it did *not* appear in the witnessed event. Thereafter, participants were asked to remember the original detail; this led to an almost complete elimination of the misinformation effect. Echterhoff, Hirst and Hussy (2005) took another approach in trying to socially discredit the misinformation (similar to Dodd and Bradshaw's procedure mentioned above but using a post-warning instead of a pre-warning) and also found a substantial reduction of the misinformation effect.

[1] This distinction between the two main types of misinformation effect is purely descriptive; it reflects the two main types of *dependent variable* in misinformation studies (i.e., what the memory assessment focuses on). It neither suggests nor forecloses any theoretical interpretations of those effects; the descriptive and the theoretical level are entirely separate. We will return to theoretical interpretations later, in our discussion section.

Generally, a considerable variety of warning procedures have been used, and perhaps not surprisingly, the results have been mixed in terms of reductions of the misinformation effect. This is precisely why we thought a more systematic approach is needed in order to find out if and to what degree warnings can safeguard against the misinformation effect. More specifically, and anticipating that the answer might not be as straightforward as implied in the last sentence, we tried to find out exactly *how effective different types of warnings are under which circumstances*. A powerful tool to answer such questions is meta-analysis.

1.1. Scope of the meta-analysis and a theoretical analysis of warnings

For both practical and theoretical reasons, we restricted our meta-analysis to *post-warning* studies. In real life – unlike in laboratory settings where researchers are aware of misleading details from a specific source of information because they have set up these conditions themselves – it is rarely possible to effectively *pre-warn* witnesses against misinformation they may *potentially* encounter at *some* point from *some* source. By contrast, we agree with other researchers (e.g. Echterhoff et al., 2005) that it would be very useful to be able to *post-warn* witnesses against misinformation, if there are good reasons to believe that they may have encountered such misinformation (e.g. from other witnesses or through the media). Even more specifically, we were interested only in post-warnings given immediately before the memory assessment, as this would be the most practically feasible timing of a warning in real eyewitness interrogations. This was the procedure in the vast majority of post-warning studies anyway; in only a handful of cases were post-warnings issued at other times (e.g. in Chambers & Zaragoza, 2001; and in some conditions in Christaansen & Ochalek, 1983; or Eakin, Schreiber & Sergent-Marshall, 2003). Focusing on post-warnings immediately before testing also resolves the difficulty of having to deal with double warnings (e.g., after the presentation of post-event information and then again before the test) and then deciding about the respective impacts of different elements of such multiple warnings.

Findings related to post-warnings before testing are also theoretically more interesting and unambiguous than findings obtained with pre-warnings or with post-warnings at earlier points in time, because the latter two are less diagnostic with respect to the processes involved. If a pre-warning resulted in a reduced misinformation effect, this could be due to enhanced attention (e.g. better scrutiny of the post-event information), enhanced remembering, or both. Similarly, post-warnings immediately after presentation of misinformation could still affect its encoding and certainly its rehearsal. By contrast, effects of post-warnings immediately before testing can only be due to an influence at the remembering (i.e., retrieval or reporting) stage. This means also that any obtained insights about the effectiveness of post-warnings have implications

with respect to the processes underlying the misinformation effect, and we will address such implications in our general discussion section.

Our analysis involved two major steps. Firstly, we developed a systematic classification of post-warning procedures along a few theoretically relevant features, in order to be able to subsequently identify effective and ineffective types of warnings.[2] This is an important step in its own right, as the diversity of warning procedures calls for some integration. The second major step was then the meta-analysis proper, which is described later. Important decisions at this stage related to inclusion and exclusion criteria and the choice of potential moderator variables to be included in

[2] For ease of expression, we will often use the term *warning(s)* only (i.e., without the *post-*). In all such cases, however, we are referring exclusively to post-warnings.

the analysis. We opted for a lenient interpretation of both the misinformation effect (i.e., we included not only 'classic' misinformation studies but also co-witness and memory conformity-type studies[3], as long as they conformed to the general misinformation design) and the term 'warning' (i.e., we also included manipulations that were not 'officially' called warnings, as long as they met our general definition), and we determined the choice of moderator variables on both theoretical grounds and on the basis of available (i.e., codable) evidence in our study set.

1.1.1. Three dimensions of post-warnings. A perhaps obvious route to a warning classification we did *not* take was to identify some sort of strength as the core dimension of warnings. The reason for this was that the strength of a warning is partly manifested in its effectiveness (in terms of limiting the misinformation effect), and therefore 'explaining' such effectiveness in terms of strength would be circular. Instead, we classified warnings along three dimensions that are conceptually unrelated to their effectiveness.

(1) *Specificity*: Warnings vary according to the degree to which they specify the incidence, source and nature of post-event misinformation. At the lowest – but perhaps also most realistic in practical terms – level of specificity (*possibility*), warnings mention only the *possible* presence of misleading or erroneous details in the post-event account. A good example is the warning used by Greene, Flynn, and Loftus (1982): "Because the police cadet was inexperienced at detailing observed crimes, some of the information in the paragraph may have been inaccurate." This level also includes procedures that can be said to *imply* the possible presence of misinformation even if this has not been explicitly stated, such as the social post-warnings used by Echterhoff, Hirst, and Hussy (2005; see below in more detail).

At the next level of specificity (*presence*), warnings positively assure that the post-event account contained misinformation but do not specify the number or nature of misleading details. A typical example is the warning used by Zaragoza and Lane (1994): "You should be aware that some of the items mentioned in the questions you answered were not in the slides you saw."

A technique developed by Jacoby and colleagues (Jacoby, Woloshyn & Kelley, 1989) and introduced to eyewitness misinformation research by Lindsay (1990) constitutes the next level of specificity (*logic of opposition*). Here, participants are instructed that the post-event account does not contain *any* accurate information relevant to the memory test, thereby setting participants' ability to correctly remember the source of a piece of post-event information *in opposition to* reporting it as an answer in the test. This type of warning is more specific than a mere *presence*-type warning, because it completely rules out the post-event account as a source of accurate information (for instance, "There is no question on this test for which the correct answer was mentioned in the story"; Lindsay, 1990).

Finally, the highest level of specificity (*identification*) involves positively identifying the misleading detail(s), as in Wright (1993): "The narrative you just read had one incorrect fact. In the slide sequence the woman did NOT have any cereal with her breakfast."

(2) *Enlightenment*: In three studies (Blank, 1998; Highhouse & Bottrill, 1995; Oeberst & Blank, 2012), the warning did not only specify *that* there was misinformation but also *why*. That is, participants were enlightened about the scientific motivation and logic of the misinformation manipulation, very similar to a good debriefing, only that this happened already within the experiment. Theoretically, this was meant to ensure an optimal internal representation of the

[3] Various terms are (not always consistently) used in the literature to refer to different forms of introducing misinformation in direct personal interaction, rather than through written (or occasionally audiotaped) narratives or questions (see Blank, Ost, Davies, Jones, Lambert & Salmon, 2013; for an overview). In *co-witness* or *social contagion* designs (e.g., Meade & Roediger, 2002), misinformation is planted by confederates of the experimenter following carefully rehearsed scripts. In memory conformity studies (e.g. Bodner, Musch & Azad, 2009; Gabbert, Memon & Allan, 2003; Wright, Self & Justice, 2000; Paterson, Kemp & Ng, 2011; Paterson, Kemp & McIntyre, 2012), pairs of witnesses produce misinformation for one another, as a natural consequence of having been shown slightly different original information. Still other ways of directly introducing misinformation involve personal questioning (e.g., Price & Connolly, 2004) or hearing rumors from others (e.g., Principe, Haines, Adkins & Guiliano, 2010).

memory task, in terms of a search for potentially two contradictory pieces of information relevant to a test question and attention to the sources of those (see Blank, 1998; Oeberst & Blank, 2012; for more background), but the procedure might also have additional motivational effects. Warnings that just state the presence of misinformation without further explanation could induce scepticism or reactance. Providing an explanation, on the other hand, may enhance participants' cooperation and motivation to show the experimenter that they can ignore the misinformation and retrieve the original information (note that while this may formally qualify as some sort of demand characteristics, these would not in and of themselves bring about correct answers unless people can actually remember the original details).

(3) *Social discrediting of the misinformation source*: Some studies use a social psychological mechanism to (implicitly) warn participants against potential misinformation, namely, discrediting the credibility of the source that provided the post-event information. This involves challenging the source's competence or their neutrality (or both). For example, in one of Echterhoff et al.'s (2005) experiments participants were told to "… 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." This warning discredits the neutrality of the witness, implying a possible intent to mislead the audience on specific details of the event. Similar warnings can be used to discredit the source's competence (see Echterhoff et al., 2005, for more background and examples). As the two discrediting strategies (challenging competence or neutrality) seem to be similarly effective (Echterhoff et al., 2005), we do not further distinguish between them.

Note that social discrediting logically presupposes that the source is identifiable as an agent and does not remain anonymous as in many studies that just provide a narrative without a specified source. Importantly, even though this means that the source of the post-event account is identifiable from the beginning, the discrediting happens only after the source has delivered the information, or it would not qualify as a post-warning for the present purposes. There are of course studies in which the credibility of the source is suspicious from the beginning, which would constitute equivalents of *pre*-warnings (e.g. Bregman & McAllister, 1982; Dodd & Bradshaw, 1980; Lampinen & Smith, 1995; Smith & Ellsworth, 1987), and are therefore not considered here.

To summarize, warnings can be distinguished along three dimensions: their specificity, the use (or not) of enlightenment, and the presence or absence of social discrediting. The first dimension affects participants' knowledge about the prevalence and nature of

misinformation, whereas the two other dimensions go beyond this purely instructional level by also affecting people's motivation to rely on the post-event account versus their own memory of the event.

1.2. General meta-analytic strategy

To identify any effects of (the three dimensions of) post-warnings on memory performance, we conducted a series of related analyses. Following our idea in the introduction that post-warning effects likely depend on the type of warning and on the specific circumstances realized in the studies, a major focus of our analyses was on moderator effects. A particular difficulty resulted from the fact that some studies explicitly compared a post-warning condition to a standard, no-warning condition, whereas others just *used* a post-warning as part of their general procedure, without focusing on the warning in its own right. The former, comparative studies are ideal for testing the principal effectiveness of warnings. Analyses of moderator effects, however, were based on the full sample of studies in order to increase the sample size and

hence the power of these analyses. This seemed particularly justified as initial analyses did not indicate any systematic differences in effect sizes (in the warning condition) between comparative and warning-only studies (see results section).

2. Method

2.1. Included studies

Initially, studies were identified through repeated Web of Knowledge searches (using "warning", "misinformation", "eyewitness suggestibility" and related search terms in various combinations). After a number of studies had been found, checking the reference lists of those papers yielded further studies. Colleagues we knew were working in this area provided additional studies. Finally, because these latter methods had located a few post-warning studies that would not have come up in literature data base searches using plausible search terms, we launched an appeal for post-warning studies through the website and mailing list of the Society for Applied Research in Memory in Cognition (SARMAC), which traditionally specializes in eyewitness memory/suggestibility research; this yielded no further studies that met our inclusion criteria. We concluded our search in May 2013.

As explained earlier, we included studies of the misinformation effect that used a single post-warning immediately before the memory test. We defined as a (post-)warning any message that explicitly stated or clearly implied (as in social post-warnings, see above) that the post-event information presented to participants had or might have contained misinformation. We did *not* count standard source monitoring instructions (or –trainings; e.g. Poole & Lindsay, 2002) as warnings, as they do not clearly state or imply that misinformation has been presented (this is also why some researchers deemed it necessary to *add* warnings to source monitoring test instructions; e.g. Lindsay, 1990; Zaragoza & Lane, 1994). Both standard misinformation and co-witness/memory conformity studies were eligible, as long as they followed the standard procedure of presenting original information followed by post-event information including misleading details (this excluded, for instance, research by Holliday and colleagues, e.g. Holliday & Hayes, 2000, in which misinformation was partly self-generated by participants). Finally, we did not include (parts of) studies that manipulated warnings *within* participants. While this certainly makes sense within the theoretical contexts of those studies, it seems rather artificial within the applied context of our meta-analysis (i.e., first let witnesses make a testimony, then warn them about misinformation, and let them tell their story again); there were only very few examples of this procedure anyway (e.g. Lindsay, Gonzales & Eso, 1995; and some studies by Holliday and colleagues).

We identified 25 published papers that met these criteria (24 journal articles and one book chapter), spanning 30 years of research but with an emphasis on more recent research (1980s: 2 papers, 1990s: 8 papers, 2000s: 8 papers, 2010s: 7 papers). We did not come across any includable unpublished work. This should be unproblematic in terms of a potential publication bias, because (1) post-warnings were mostly not the primary focus of the included papers and (2) both the absence and presence of post-warning effects would be equally newsworthy; therefore, there is no reason to assume a selection bias on the basis of any effects in one or the other direction. Also, in our judgment[4], the 25 included papers are reasonably representative of misinformation research in general, in terms of the covered mix and range of study features (see below, *study characteristics*).

[4] It is hard to say how representative the studies covered in our meta-analysis are relative to 'the population', that is, the universe of (real-life) situations where misinformation biases memory reports, for the simple reason that no one has ever defined this universe. Nor do we believe that such a universe can ever be exhaustively described (this is an instance of the *induction problem* discussed by the Vienna Circle logical positivist philosophers, and famously declared unsolvable by Popper, 1934/1959). That is, our included studies are probably as representative of 'the population' as misinformation research in general is, no more and no less.

2.2. Coded information

2.2.1. Effect sizes. Depending on the memory assessment(s) and the number of included experiments/conditions, a given paper contributed between 2 and 16 effect sizes (see Appendix). In line with the distinction between the two major manifestations of the misinformation effect, we computed separate effect sizes for impaired memory performance for original details (*Original Memory*) and for *Misinformation Endorsement*. Conceivably, post-warnings might be differentially effective for these two types of effects, and therefore we analysed them separately throughout.

The unit for coding effect sizes was a whole study (if it contained only one experiment), an experiment within a study, or a condition within an experiment, as long as the latter was varied *between* participants (most importantly, a post-warning vs. no warning manipulation, but also various others depending on the nature of the research); the guiding principle was to have separate effect sizes for separate groups of participants. The one exception to this rule was that the Original Memory and Misinformation Endorsement effect sizes by necessity came from the same participants if both measures had been assessed in a study/experiment/condition (but then the two types of effect sizes were analysed separately anyway). Also, in a few studies (see Appendix) the same control group was compared to a misled and a misled-postwarned group, or to two groups involving different types of warning, thereby making the involved effect sizes dependent to some degree. While this is not ideal in terms of statistical integration, we opted to retain these effect sizes because they still represent valuable information (and also they can be excluded for specific analyses if necessary, see below). Finally, in a few cases where more than one memory test had been used to assess memory performance after a post-warning (or at a comparable time in no-warning conditions), we coded only the results from the first memory test (or if these were not reported, the temporally closest test results that were reported), again to avoid dependencies in the data as far as possible, and also because earlier tests may have influenced memory performance on later tests in ways unrelated to the warning.

[80]

---

[81]

Altogether, this yielded 59 Original Memory effect sizes (40 post-warning, 19 no-warning, i.e. offering 19 direct post-warning vs. no-warning comparisons) and 96 Misinformation Endorsement effect sizes (53 post-warning, 43 no-warning, i.e. 43 direct comparisons). We calculated odds ratios as effect sizes, the recommended type of effect size for data based on inherently binary responses (i.e. correct vs. incorrect; Haddock, Rindskopf & Shadish, 1998). Specifically, the odds of producing a correct vs. incorrect response in the misled condition (i.e. with misinformation provided) were compared to the respective odds in the control condition, and the odds ratios for Original Memory and Misinformation Endorsement were always computed such that higher odds ratios represent stronger misinformation effects. To illustrate, a study assessing Original Memory may have found 80% correct (and therefore 20% incorrect) responses in the control condition (translating into odds of 4.0 – 80% divided by 20%) and 60% correct in the misled condition (translating into odds of 1.5 = 60%/40%); the resulting odds ratio is 4.0/1.5 = 2.67. Odds ratios higher than 3.0 are considered large effect sizes (Haddock et al., 1998), and an odds ratio of 1.0 represents no difference between conditions (i.e., no misinformation effect).

Due to the enormous heterogeneity of the experimental designs and conditions used in the 25 papers, we had to make decisions about which conditions to use, drop, combine, etc., in order to be able to calculate the most meaningful effect size(s). Also, a few of the Misinformation Endorsement effect size calculations were tricky in that there were performances of 0% in the control condition (which makes it impossible to calculate an odds ratio for inclusion in the meta-analysis). In such cases, we combined zero and nonzero performances, or replaced zero values with non-zero values

from comparable control conditions on a case-by-case basis. These data treatments are too idiosyncratic to be reported here; we explain them in great detail in the notes to the Appendix. Finally, we note that when a two-alternative forced-choice (between original and misleading details) recognition test is used to assess memory performance, this can be classified under both Original Memory and Misinformation Endorsement (because not choosing the original detail automatically means choosing the misleading detail). There were nine such cases in the data set (6 post-warning, 3 no-warning), and we chose to code them both ways, as they do have both aspects and to not lose information.

Table 1

Study characteristics, coding categories, number of papers (in bold print) falling into each category, and outcomes of moderator analyses for memory for original details (Original Memory, OM) and Misinformation Endorsement (ME)

Study characteristic	**Categories and number of papers**	**OM**	**ME**
Type of participants	Students: **22**, Children: **3**	ns	s/ns*
Misinformation paradigm	Standard: **21**, Direct personal[1]: **6**	ns	s/ns*
Delay of misinformation (after original event)	0-15 min: **16**, 20-60 min: **6**, 1-4 days: **4**	ns	ns
No. of misleading details	1-2: **8**, 3-4: **9**, 5-8: **7**, >8: **4**	ns	s/ns*
Type of misleading details	Contradictory: **17**, Supplementary: **11**	ns	ns
Delay of test (after misinformation)	0-15 min: **17**, 20-60 min: **3**, 1-4 days: **5**, 1 week: **4**, 5 weeks: **1**	ns	s/ns*
Type of test	Recall: **11**, Recognition[2]: **18**	ns	s*
Specificity of warning	Possibility: **7**, Presence: **13**, Logic of opposition: **5**, Identification: **2**	ns	s/ns*
Enlightenment	Yes: **3**, No: **22**	p = .01	p = .02
Social discrediting	Yes: **5**, No: **22**	ns	ns

Note. [1] 'Direct personal' includes different ways of introducing misinformation in direct personal interaction, as in co-witness or in memory conformity studies or through personal questioning or through personally transmitted rumors (cf. footnote 2). [2] See the Appendix for a further breakdown of recognition tests. s/ns***** = the initially significant moderator effect disappears if four extreme effect sizes stemming from one study (Principe et al., 2010) are removed. s* = the moderator effect remains significant but is strongly reduced if four extreme effect sizes stemming from one study (Principe et al., 2010) are removed; see text for details.

2.2.2. Study characteristics. For later use in moderator analyses, we coded a number of study characteristics, some of which were of core theoretical interest (specifically, the three dimensions of warnings as described above). Other characteristics were partly also chosen because of theoretical considerations; as these did not translate into significant moderator effects, however (see results), we do not detail those ideas here. The coded study characteristics along with the respective coding categories and numbers of papers that fell into each category are given in Table 1 (ignore the two last columns for the moment), in order to convey a rough impression of the post-warning 'research landscape' (note that

some multi-experiment/condition papers could fall into more than one category, so that the numbers do not necessarily add up to the total number of papers; more detailed information about the number of *effect sizes* at the category levels will be given later along with the moderator analyses). This post-warning studies landscape was fairly heterogeneous in terms of the general (i.e., not warning-related) study characteristics, but always within a typical range, and is in this sense reflects the breadth of misinformation research in general. All codings of study characteristics were double-checked at several times between the two present authors, and any inconsistencies resolved by discussion. Most of the coded features are fairly unambiguous and easily verified in any case, leaving little room for error.

3. Results

All statistical analyses followed the standard procedures and recommendations in Haddock et al. (1998) and Hedges and Olkin (1985), using a fixed effects model[5] and inverse effect size variance weights (essentially, weighting for sample size). Also, we used not the odds ratios themselves but the respective log odds ratios (i.e., the natural logarithms of the odds ratios; all log odds ratios are given in the Appendix) for the actual analyses, as this is mathematically easier. Because the odds ratios are more intuitive, however, we report all results as re-transformed into this format.

All analyses were conducted in a standard and a conservative version, the difference lying in the weights assigned to effect sizes calculated from between- and within-participant misinformation manipulations. In within-participants designs, the same participants contribute data points to both the misled and control conditions. Although the dependencies between memory performances for individual items are typically not strong, treating them as independent (i.e. as if they had been obtained in a between-participants design) is perhaps slightly optimistic. As a control for this, our conservative and perhaps too 'pessimistic' analysis treated the observations as if they were completely *de*pendent,

effectively halving the weights for those effects sizes. It turned out that the results of the standard and conservative analyses did not differ much and led to the same conclusions. Therefore, we report the standard analyses by default, but do report the conservative ones as well in the rare cases in which they differed.

3.1. Descriptive and preliminary analyses

Figures 1 and 2 show stem-and-leaf plots of post-warning and no-warning misinformation effect sizes, for Original Memory and Misinformation Endorsement, respectively. In both cases, the difference between post-warning and no-warning effect sizes is visible to the naked eye. Note that the stem width in Figure 2 is wider, reflecting the generally larger effects sizes for Misinformation Endorsement. Before we can follow this up in a moderator analysis, however, we need to briefly address one potential difficulty with the Original Memory effect sizes in particular: Different studies/experiments/conditions used different types of control conditions to assess the misinformation effect, and this might conceivably affect the effect sizes. Specifically, some studies used a control condition in which the original event detail (to be contradicted in the misled condition) was repeated in the post-event information. In other cases, the original event details was only referred to generically in the post-event information (e.g., when the original detail was a hammer, a *tool* was mentioned), and in still others the critical detail was not mentioned at all. Now, theoretically, repeating the original detail should strengthen control performance through rehearsal, while no such improvement can be expected

[5] A sometimes recommended alternative to a fixed effects model is a random effects model, which assumes that the population effect size randomly varies around the estimated value. We do not think that this is a theoretically sensible statement and therefore opted for the fixed model, which is also simpler and easier to understand.

when the detail is not mentioned; the generic control condition type should be somewhere in between. To check for any such effects, we conducted two separate moderator analyses (for the post-warning and no-warning effect sizes) with the three control condition types plus a fourth, mixed category (when more than one type had been used) as levels of the moderator variable. Although the trend was in the expected direction in both cases (ignoring the mixed category), the moderator effects were not significant (*p*s > .25). We conclude therefore that if type of control condition has an influence, it is not a strong one. Therefore, we ignored this distinction for all subsequent analyses and treated the effect sizes obtained with different control conditions alike.

Post-warning (*n* = 40)		**No warning (*n* = 19)**
99988875	*0*	*9*
99997655443332210000	1	0
5100000	2	11455778
8511	3	559
	4	
	5	29
	6	45
1 ES	7+	2 ESs

Figure 1: Stem-and-leaf plot of misinformation effect sizes (Original Memory) in post-warning and no-warning studies/experiments/conditions. Effect sizes (ESs) are odds ratios. The stem width is 1 (e.g., a stem of 1 and a leaf of 4 translates into an odds ratio of 1.4). ESs < 1 (in italics) represent inverse misinformation effects (i.e. better misled than control performance).

Post-warning (*n* = 53)		**No warning (*n* = 43)**
1111111111111111111*10000*	0	*0*11
7644443333332222222	0	222223333333344567788
66331	10	1123366666788
1	20	37
4 ESs	>50	4 ESs

Figure 2: Stem-and-leaf plot of misinformation effect sizes (in terms misinformation endorsement) in post-warning and no-warning studies/experiments/conditions. Effect sizes (ESs) are odds ratios. Different from Figure 1, the stem width is 10 (e.g., a stem of 10 and a leaf of 6 translates into an odds ratio of 16). ESs < 1 (in italics) represent inverse misinformation effects (i.e. better misled than control performance).

3.2. Main analyses I: Do post-warnings reduce the misinformation effect?

To answer the basic question if and to what degree post-warnings reduce the misinformation effect, we started out with comparing (separately for Original Memory and Misinformation Endorsement) all post-warning effects sizes to all no-warning effect sizes (as shown in Figures 1 and 2). The respective moderator analysis (i.e. using post-warning vs. no-warning as a moderator) returned a highly significant moderator effect for both Original Memory, $Q(1)$[6] = 20.96, p < .001, and Misinformation Endorsement, $Q(1) = 27.40$, $p < .001$. The associated post-warning and no-warning effect size estimates (weighted averages across studies, i.e. taking sample size into account; see Haddock et al., 1998; Hedges &

[6] *Q* is the test statistic for moderator effects, that is, differences between two (or more) meta-analytic sets of effect sizes (in this case, from post-warning and no-warning studies), analogous to an *F* value in an ANOVA.

Olkin, 1985) were 1.49 (95% CI: 1.29, 1.73) and 3.40 (95% CI: 2.47, 4.68) for Original Memory, and 2.08 (95% CI: 1.69, 2.55) and 4.84 (95% CI: 3.80, 6.16) for Misinformation Endorsement.

As with most moderator analyses, one limitation of the preceding analyses is that they are essentially correlational: Studies with post-warnings show smaller misinformation effects than studies without post-warnings, but this does not strictly mean that the difference was caused by the post-warning. Fortunately, as post-warnings were experimentally manipulated in a subset of the studies/experiments/conditions, we are in a much better position here: By focusing only on direct comparisons between post-warning and no-warning conditions, any resulting moderator effects can be causally attributed to the post-warning manipulation. There were 19 such direct comparisons for Original Memory and 43 for Misinformation Endorsement. The means and 95% confidence intervals resulting from the respective moderator analyses are visually displayed in Figure 3. The no-warning effect sizes are by necessity the same as in the previous analysis (as the same studies had been included before), and the post-warning effect sizes – 1.47 (95% CI: 1.06, 2.05) for Original Memory and 2.10 (95% CI: 1.63, 2.69) for Misinformation Endorsement – are very similar

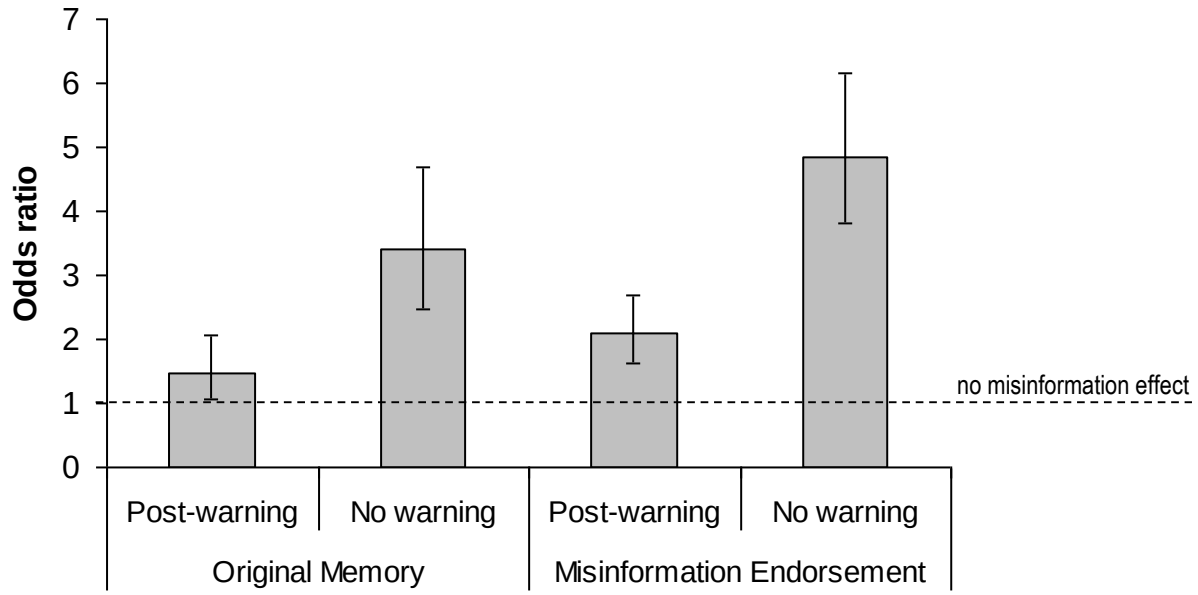

Figure 3: Misinformation effect sizes (odds ratios) as a function of post-warning. Error bars represent 95% confidence intervals.

to those obtained from the full set of post-warning studies. Not surprisingly, then, the moderator effects were again highly significant, $Q(1) = 18.12$, $p < .001$ for Original Memory, and $Q(1) = 23.31$, $p < .001$ for Misinformation Endorsement. But this time we can draw a much stronger, causal conclusion: Post-warnings do indeed reduce the misinformation effect. We can also express the amount of this reduction: Across our sample of direct comparisons, post-warnings reduced the Original Memory misinformation effect to 43% (= 1.47/3.40) of its original (i.e. no-warning) size. Incidentally, the exact same reduction (43% = 2.10/4.84) emerged for the Misinformation Endorsement effect.

Still, it is worth noting that neither the Original Memory nor the Misinformation Endorsement effect were fully eliminated across our sample of comparisons, as can also be seen from the 95% confidence intervals in Figure 3, which

do not include the no-effect value of 1 (although the Original Memory post-warning CI comes close). This conceivably reflects the heterogeneity of the post-warnings, some of which may be more and others less effective. Subsequent moderator analyses will address such differences. Note that, for power reasons, these will be conducted on the basis of the full set of post-warning studies. An important precondition for this to be possible is that the full set is comparable to the subset for which we have just established the causal effectiveness of post-warnings. Fortunately, as the near-identical results between the two sets of studies show, this can be safely assumed (it is also confirmed in moderator analyses that directly compare the post-warning Original Memory and Misinformation Endorsement effects obtained in comparison studies to those obtained in warning-only studies; both *p*s > .90).

3.3. Main analyses II: Does the effectiveness of post-warnings depend on study characteristics and the nature of the warnings?

3.3.1. Moderator effects of general study characteristics. Does the effectiveness of warnings depend on the circumstances? The next analyses, which we report only very briefly, checked if the effectiveness of post-warning (or, as a proxy, the remaining size of the misinformation effect after the post-warning) was associated with a number of general study characteristics mentioned earlier (i.e., not including the three specific dimensions of warnings; those will be treated separately below). The results of the respective moderator analyses are shown in the upper part of Table 1. They are easily summarised: Type of participants, misinformation paradigm, delay of misinformation, number and type of misleading details, as well as delay and type of test, had no significant links whatsoever with Original Memory. There were some initially significant links of these moderator variables to Misinformation Endorsement (see Table 1), but most of these disappeared after four unusually large effect sizes (i.e. remaining Misinformation Endorsement effects) contributed by one single study (Principe et al., 2010) were removed. This study differs from other included studies in that it introduced misinformation in a rather untypical way, by having children overhear a conversation between adults, or encounter it as rumour from other children; recall of the overheard/rumoured misinformation was assessed one week later. This produced very strong misinformation effects, and even though the effects were reduced after post-warning, they were still strong enough to bias some of the moderator analyses. (Alternatively, we might have excluded this study from the outset, but there was no compelling theoretical reason and we did not want to lose too many data points.)

The one moderator effect (for Misinformation Endorsement) that remained significant even after excluding the Principe et al. effect sizes was that of type of memory test (recall vs. recognition). The initial moderator effect was $Q(1) = 14.94$, $p < .001$, with a stronger remaining Misinformation Endorsement when a recall test was used (odds ratio = 4.11; 95% CI: 2.75, 6.14) rather than a recognition test (odds ratio = 1.64; 95% CI: 1.29, 2.08). As Principe et al. had used a recall test, the recall odds ratio reduced to 3.21 (95% CI: 2.12, 4.87) after the four effect sizes from this study had been removed, and the recall vs. recognition moderator effect was reduced to $Q(1) = 7.56$, $p = .006$. That is, in either analysis, post-warnings were less effective, in terms of curbing Misinformation Endorsement, when recall tests were used. We do no want to make too much of this isolated finding – after all, given the considerable number of moderator analyses conducted (14), it might be a type I error. Generally, the message from these moderator analyses so far is that the effectiveness of post-warnings does not seem to be linked to the particular circumstances of the investigated studies.

3.3.2. Specific effects of warning dimensions. In our next set of analyses, we explored differences in warning effectiveness linked to the three warning dimensions identified earlier – specificity, enlightenment, and social discrediting. The results of these analyses are summarised in Table 1 (lower part) and in Figures 4 to 7. Warning specificity (Figures 4 and 5) did not significantly moderate the misinformation effect in terms of Original Memory, $Q(3) = 3.83$, $p = .28$, but initially did in terms of Misinformation Endorsement, $Q(3) = 11.13$, $p = .011$. As in previous

moderator analyses, however, this effect was owed to the four Principe et al. effect sizes that, in this analysis, fell under the logic of opposition category (which also explains the surprisingly large surviving Misinformation Endorsement effect at this moderator level). After removal of these effect sizes, the moderator effect was reduced to $Q(3) = 5.30$, $p = .15$. There was a similar seeming anomaly for Original Memory, where perhaps counterintuitively a moderately strong misinformation effect survived at the most specific level, identification. This was owed to seven effect sizes coming from one study, Eakin et al.

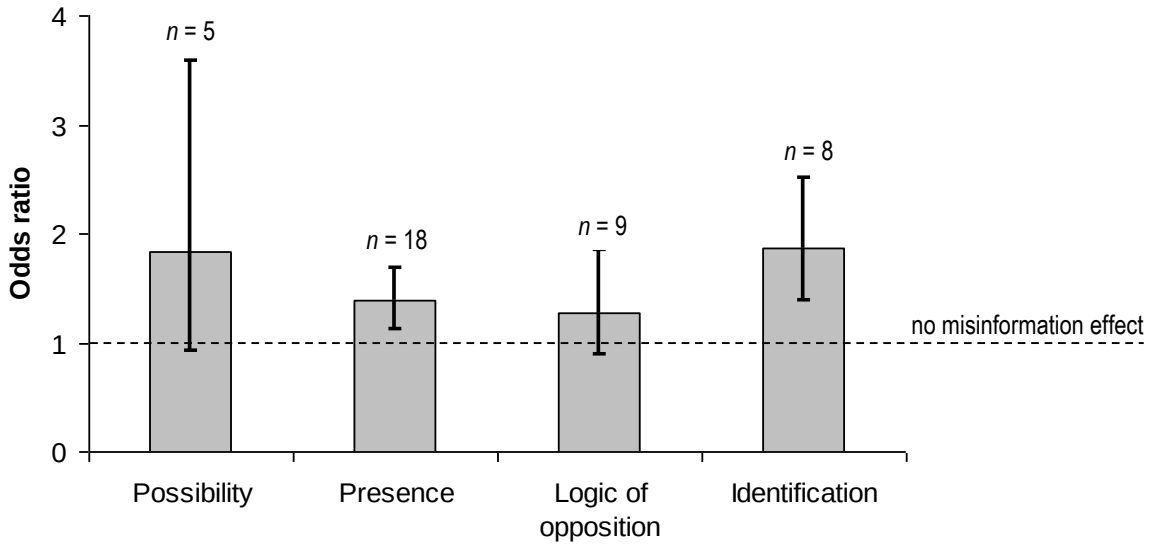

Figure 4: Misinformation effect sizes (odds ratios; **Original Memory**) as a function of warning specificity. Error bars represent 95% confidence intervals. *N*s denote the number of effect sizes at each moderator level.

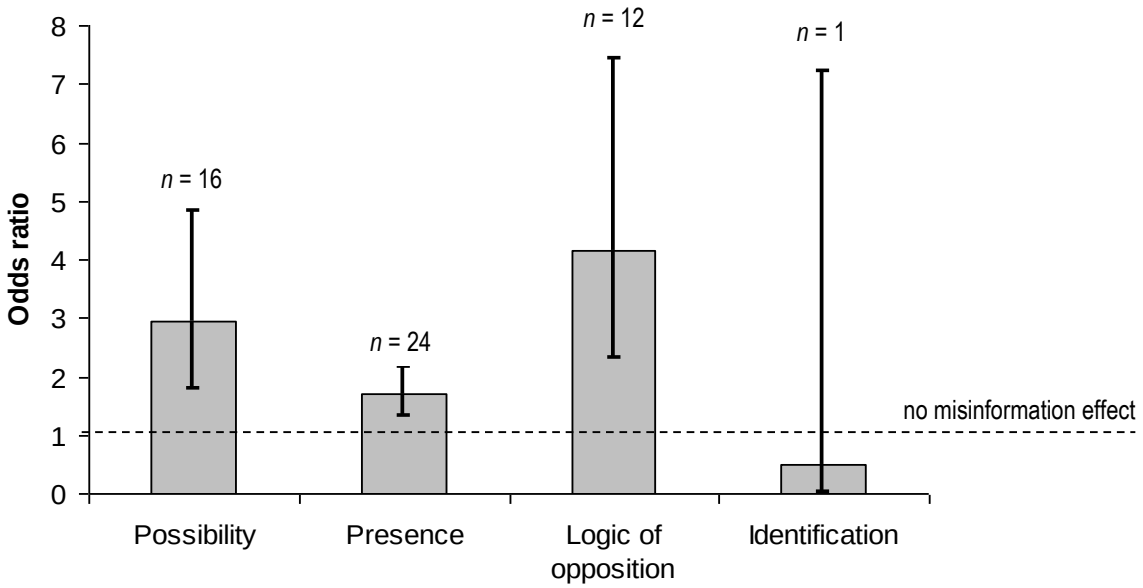

Figure 5: Misinformation effect sizes (odds ratios; **Misinformation Endorsement**) as a function of warning specificity. Error bars represent 95% confidence intervals. *N*s denote the number of effect sizes at each moderator level.

[84]

(2003), and these authors explain the persistence of a misinformation effect through a mechanism featured and facilitated in their study, retrieval blocking. That is, both the Original Memory and Misinformation Endorsement results patterns are somewhat disproportionally influenced by idiosyncratic features of individual studies.

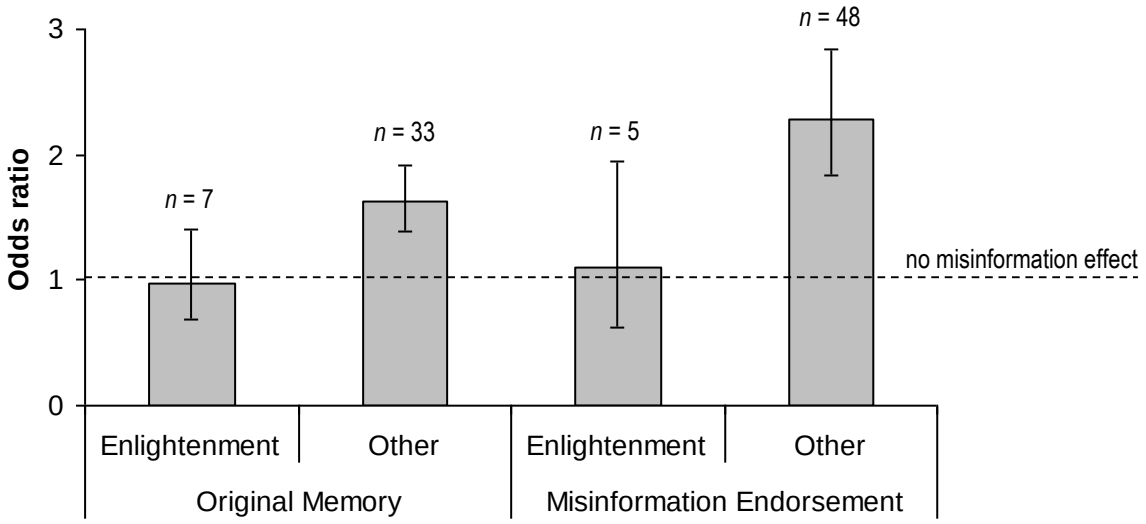

Figure 6: Misinformation effect sizes (odds ratios) as a function of enlightenment. Error bars represent 95% confidence intervals. *N*s denote the number of effect sizes at each moderator level.

Enlightenment (Figure 6; used in Blank, 1998; Highhouse & Bottrill, 1995; Oeberst & Blank, 2012) significantly moderated both Original Memory and Misinformation Endorsement effects, $Q(1) = 6.21$, $p = .013$ and $Q(1) = 5.54$, $p = .019$, respectively. In both cases, there was no significant misinformation effect left after enlightenment, as can be seen from the confidence intervals in Figure 6. We note, however, that the moderator effects were only marginally significant in our conservative analyses using 'pessimistic' weights for within-participants misinformation manipulations, $Q(1) = 3.40$, $p = .07$ and $Q(1) = 3.16$, $p = .08$, respectively.

Finally, and perhaps surprisingly, social discrediting (Figure 7; used in Echterhoff et al., 2005, 2007; Greene et al., 1982; Paterson, Kemp, & McIntyre, 2012; Price & Connolly, 2004) was essentially unrelated to the effectiveness of warnings, $Q(1) = 0.08$, $p = .79$ for Original Memory, and $Q(1) = 0.00$, $p = .99$ for Misinformation Endorsement. There were some studies with very small misinformation effects after source discrediting (Echterhoff et al., 2005, 2007; Price & Connolly, 2004) but also others with stronger remaining effects (Greene et al., 1982; Paterson et al., 2012).

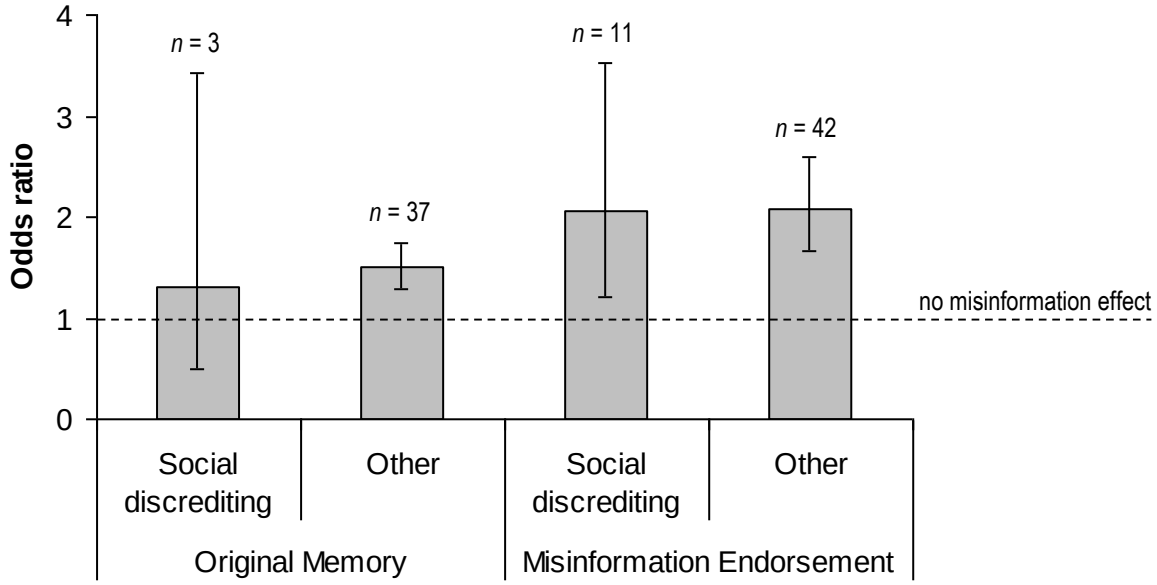

Figure 7: Misinformation effect sizes (odds ratios) as a function of social discrediting. Error bars represent 95% confidence intervals. *N*s denote the number of effect sizes at each moderator level.

3.4. Main analyses III: A breakdown of processes involved in post-warning effects

So far, we have been looking at post-warning effects in a global perspective, not focusing too much on *how* these effects are achieved exactly. One step in this direction is to look at misled and control performances separately: Principally, keeping in mind that the misinformation effect is defined as the difference between memory performance in the misled and control conditions of a misinformation design, post-warnings can reduce the misinformation effect by either improving misled performance (relative to a no-warning condition) or by reducing control performance. While the former is probably more intuitive (and certainly more desirable from an applied point of view), the latter possibility cannot be ruled out a priori, at least not with respect to memory for original details.[7]

The argument for the latter possibility would be that post-warnings not only alert participants to the (possible) presence of misleading details in the post-event information but, beyond this, may make participants more sceptical of any information that is familiar and cannot be unequivocally attributed to the original information only (i.e., declared as 'safe'). Importantly, this may apply to items in the control condition as well, foremost to control items that have been repeated in the post-event information (remember our earlier discussion of different types of control items) but also to items referred to in neutral terms and even to items that were not repeated but the source of which has been forgotten (and therefore only familiarity remains). The idea is very similar to the 'tainted truth' argument put forward by Echterhoff et al. (2007; see also Szpitalak & Polczyk, 2010, 2011), only that these authors focus more on the recall of non-critical correct information (i.e., what would be 'filler' or general event items in a misinformation design).

In short, any benefits of post-warnings in terms of improved memory performance for misled items (by whichever mechanism; we return to this question in our discussion section) may be accompanied by costs in terms of

[7] This analysis does not apply to Misinformation Endorsement, because the general memory scepticism argument we develop would predict better control performance (i.e., less Misinformation Endorsement), not worse performance here.

reduced performance for control items (due to general scepticism towards potential misinformation). Fortunately, these two effects/possibilities can be disentangled by looking at the experimental and control performances in post-warning and no-warning conditions separately. To illustrate, if misled memory performance rose to 70% correct after post-warning, compared to 60% correct without warning, this would constitute a Δ = 10% misled performance benefit. Correspondingly, if control performance dropped to 75% correct after post-warning, relative to 80% correct without warning, this would show a Δ = 5% control performance cost. Altogether, the initial (i.e. no-warning) misinformation effect of Δ = 20% (80% - 60%) would have been reduced to Δ = 5% (75% - 70%), but this reduction would be a combination of a 10% misled benefit and a 5% control cost.

To assess the relative importance of these two effects for post-warning, we calculated and meta-analysed odds ratios as before, but this time not odds ratios representing misinformation effects (i.e., comparing misled and control performances) but odds ratios representing misled benefits and control costs as introduced above (i.e., the Δ = 10% misled benefit in the above example would translate into an [70%/30%]/[60%/40%] = 2.33/1.50 = 1.56 odds ratio, etc.). Additionally, we calculated meta-analytic performance averages (i.e., weighted averages of the performances in individual studies/experiments/conditions, using the same weights as before in the meta-analyses of the odds ratios), in order to give a more tangible impression of how post-warning affects misled and control memory performance. Note that these analyses can only be done on the basis of studies that directly compare post-warning and no-warning conditions (which has the additional advantage of allowing causal conclusions, as in the direct comparisons in our main results section I). Furthermore, the assessment of control costs requires separate post-warning and no-warning control conditions, which excludes a few studies that used just one common

control condition (to be compared to post-warning and no-warning misled conditions). This left 16 effect sizes (from 8 papers) for our analyses.

Table 2

Meta-analytic performance averages (% correct memory for original details) and performance differences reflecting benefits and costs of post-warnings in the misled and control conditions of misinformation designs

Post-warning		No warning	
Misled	Control	Misled	Control
63.3	71.4	51.5	74.7
[---------- Misled benefit: Δ = 11.8 ----------]			
		[---------- Control cost: Δ = 3.3 ----------]	

Table 2 shows the relevant meta-analytic performance averages and differences. Descriptively, the misled benefits are 3.5 (= 11.8/3.3) times larger than the control costs. The outcomes of the meta-analyses using the misled benefits and control costs odds ratios (as explained above) are visualised in Figure 8. The misled benefits odds ratio of

1.74 (representing 74% better odds of obtaining correct information in the misled condition after post-warning) is significantly above the no-effect ratio of 1 (as can be seen from the confidence intervals in Figure 8). By comparison, the 1.21 control costs odds ratio is much lower and does not differ significantly from the no-effect value of 1. In short, there is clear meta-analytic evidence that post-warnings lead to a real improvement of memory performance for original details that have been the target of misinformation, whereas the evidence for associated costs to control memory performance is more tenuous.

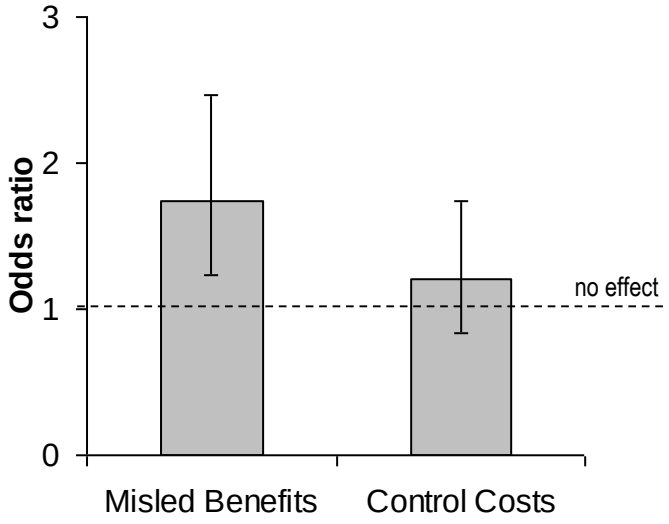

Figure 8: Effect sizes (odds ratios) representing memory performance benefits and costs of post-warnings in the misled and control conditions of misinformation designs. Error bars represent 95% confidence intervals.

For completeness, we also briefly report the meta-analytic performance averages for Misinformation Endorsement. Across 25 direct comparisons (stemming from 10 papers), misled and control Misinformation Endorsement in the no-warning condition was 42.3% and 17.7%, respectively. The corresponding post-warning figures are 28.5% and 19.5%. As an aside, we can derive from these figures (and the corresponding ones for Original Memory), another way of expressing the reduction of the misinformation effect after post-warning, which is perhaps more intuitive than the measure derived from the odds ratios earlier (in the main results I section): The initial performance difference of 42.3% - 17.7% = 24.6% was reduced to 28.5% - 19.5% = 9.1% (rounding error) after post-warning. The corresponding reduction in terms of Original Memory (cf. Table 2) was from an initial 23.2% difference to an 8.1% difference after post-warning. The two reductions are remarkably similar, leaving just over a third of the initial effect in both cases.

4. Discussion

The main message of this meta-analysis is that post-warnings are surprisingly effective, reducing the misinformation effect to somewhere between a third and half of its 'normal' size (i.e., without warning), depending on the exact way of measuring this reduction (i.e., on the basis of odds ratios or performance differences). This holds about equally for both core measures of the misinformation effect (cf. Higham, 1998; Pansky, Tenenboim & Bar, 2011), impaired memory performance for original event details and misinformation endorsement. Further, with respect to original event memory, the benefits of post-warnings in terms of improving misled performance clearly outweigh any costs in terms of reduced

control performance. We also emphasize that, as analyses of subsets of studies with direct experimental comparisons of post-warning and no-warning studies have shown, the reductions of the misinformation effect are in fact *caused* by the post-warnings.

The range of post-warning effects was considerable, though: In some cases post-warnings were completely ineffective, and in others the misinformation effect was completely eliminated. Moderator analyses sought to identify features of studies and post-warnings that might explain some of this variability (using the remaining size of the misinformation effect as a proxy for the effectiveness of the post-warning), but yielded only two hints. (1) Stronger misinformation endorsement remained after post-warnings when recall tests, as opposed to recognition tests, were used. We are not sure what this means, particularly as recognition tests are a quite heterogeneous category, comprising yes/no tests, forced-choice tests but also source monitoring tests (more fine-grained analyses – not reported in detail in our results section – did not yield a more meaningful picture, though). (2) Post-warnings using an element of enlightenment (i.e., information about the background, purpose and design of the study, similar to a good debriefing but already within the study; see e.g. Oeberst & Blank, 2012) worked much better than other types of post-warning and completely eliminated the misinformation effect. Other dimensions of warnings (specificity and social discrediting) were less influential.

A general problem with these moderator analyses, however, was that they were partly based on very small numbers of studies/effect sizes at particular moderator levels, and therefore vulnerable to effects of idiosyncratic features of studies. Also, due to such small numbers, some of the moderator analyses probably had insufficient power. We think it is entirely possible that if this meta-analysis were done again in ten years, with many more studies included, the moderator analyses might reveal a more interesting picture. Some of the moderators seem to make theoretical sense. For instance, with respect to warning specificity, one should expect warnings that mention only the possibility of misinformation, rather than positively assert its presence, to be less effective. Descriptively, this was indeed so (when outlying effect sizes were excluded), but the moderator effect did not reach significance. At present, therefore, we can only draw very limited and tentative conclusions from the results of these moderator analyses (specifically, post-warnings containing an element of enlightenment seem promising candidates, and the type of memory test used might be important as well). Clearly, more research would help here.

4.1. Theoretical and practical implications

So far we have been silent as to the theoretical processes involved in the operation of post-warnings, as this was not essential for investigating their effectiveness. Furthermore, apart from the hints provided from our costs-and-benefits analysis above, there is little conclusive evidence to be gained here from our meta-analysis that (necessarily) focussed on overall performance effects. Still, a

few comments are in order. Principally, post-warnings can be effective through undermining any of the processes that otherwise (i.e., in the absence of a post-warning) are supposed to lead to a misinformation effect. Very roughly (see much more detailed coverage by e.g. Belli & Loftus, 1996; Loftus, 1991; McCloskey & Zaragoza, 1985; Zaragoza, Belli & Payment, 2006), three major mechanisms can be distinguished: (1) temporary or permanent memory impairment, that is, impaired ability to remember event details that were the target of misinformation, (2) biased responding in favor of the misinformation, at the expense of reporting the original event details, and (3) source

misattribution, that is, misattributing suggested details to the original event (which in itself can be a consequence of different processes; Lindsay, 2008). These mechanisms can operate in isolation or in combination. A fourth explanation that cuts across the three just mentioned highlights the importance of *memory conversion* processes; it blames the misinformation effect on suboptimal use of original memory information, due to an inadequate representation of the memory task (Blank, 1998; Oeberst & Blank, 2012; see also Lane, Roussel, Villa, & Morita, 2007, for a related approach).

In line with these theoretical ideas, post-warnings may then reduce the misinformation effect either by removing (temporary) memory impairment, by undermining response biases, by improving source discrimination, or by providing a more adequate task representation (with expected consequences in terms of the previous three mechanisms, i.e., replacing a memory search-and-accept strategy with a memory search-and-discriminate strategy; Oeberst & Blank, 2012). It is impossible to determine, on the basis of the meta-analytic data, how much each of these mechanisms contributes to the overall post-warning effects; this is often also not clear within the original studies.

The relative contributions probably depend on specific aspects of the studies, for instance on the type of memory test used. For instance, in a standard two-alternative forced-choice recognition test (e.g., Loftus et al., 1978; four of the 40 Original Memory effect sizes came from such tests), de-biasing may be entirely sufficient to restore original memory performance, as discrediting the misleading detail automatically favours reporting the original detail. By contrast, in recall tests (accounting for ~ 40% of the Original Memory effect sizes) de-biasing in itself may not help very much unless access to – or discrimination of – original event details is otherwise facilitated or restored. The fact that there was no big difference between the Original Memory post-warning effect sizes for recall and recognition tests could mean that de-biasing was indeed supplemented by other post-warning mechanisms in studies using recall tests. As mentioned earlier, the current sample size was too small for more refined analyses along these lines, but it would be worthwhile to investigate the effectiveness of particular warnings in combination with particular memory tests in future studies or meta-analyses.

Further, different types of warning probably rely on different warning mechanisms to differing degrees. Social post-warnings, in undermining the credibility of the misinformation source, essentially rely on de-biasing; identification-type warnings achieve the same through naming the misleading detail(s). Logic of opposition-type warnings enhance source discrimination, and enlightenment aims to change the internal representation of the memory task, with consequences for potentially all three major mechanisms, de-biasing, source discrimination, and memory search (which is perhaps why it is so effective). But the latter holds for all types of warning to some degree: After de-biasing, for instance, the participants need to find a different answer to the test question, which necessarily affords increased attention to sources and additional memory search efforts. In practice, therefore, all of the key warning mechanisms are likely to be involved to various degrees in all types of post-warnings.

In any case, a very general theoretical conclusion from this research points to the malleability of the misinformation effect: Whether or not (or to what degree) people's rememberings are influenced by post-event misinformation depends on the conditions of remembering (cf. the theoretical analysis in Oeberst & Blank, 2012, pp. 154-155). One important condition, as we have seen, is the presence or absence (and likely the nature) of post-warnings, but there are others. For instance, Bekerian and Bowers (1983; Bowers & Bekerian, 1984) eliminated the misinformation effect by reinstating the original encoding context at test. Similarly, Lindsay and Johnson (1989a) did not find a misinformation effect when they used a source monitoring test instead of a forced-choice recognition test.

Generally, this points to the importance of *memory conversion* (Tulving, 1983) with respect to suggestibility effects in remembering: What is remembered in a given situation depends not only on what (presumably) is or is not in memory, but also on exactly how this information is used in the testimony, in the light of additional information provided in the test situation/the social context (see Blank, 1998, 2005, 2009; Oeberst & Blank, 2012; for more detailed analyses). Compared to what we know about the 'hard' memory processes involved in encoding, storage, and retrieval, we know very little about these 'soft' memory conversion processes. A better understanding of these might also help to develop more efficient post-warning techniques.

From a practical point of view, many of the post-warnings used in our featured studies are unsatisfactory, in that they relied on positively asserting the presence of misinformation, or even pointing out the misleading details. This is of course possible, and makes sense, in laboratory studies where the investigators know about the misinformation because they have planted it themselves. In applied settings, however, it will typically be uncertain whether and what kind of misinformation has been provided (e.g. by other witnesses or by the media), and in the rare cases where the misinformation is known, there would not necessarily be a need for a post-warning – precisely because the misinformation is known and therefore would be recognized anyway if erroneously produced by a witness. What would be needed, then, is more research into the development of effective post-warnings that work under conditions of uncertainty about the presence, nature and extent of misinformation. Oeberst and Blank (2012, pp. 155-156) provide some ideas, but there is certainly a lot more that could be done. The encouraging findings from this meta-analysis (specifically, post-warnings mentioning the possibility of misinformation only were not significantly less effective than others) suggest that research systematically directed at developing effective realistic warnings may be worth the while and effort.

Finally, we speculate if the post-warning research reviewed here can be generalized (in some shape and form) to other settings where misinformation effects have been observed. While the eyewitness misinformation effect typically refers to a situation where misinformation is provided *after* a focal event (but see Lindsay & Johnson, 1989b, for a rare exception), other misinformation effects (reviewed by Lewandowsky, Ecker, Seifert, Schwarz & Cook, 2012) pertain to situations and research settings where some misinformation is contained already in an initial account[8] and then turns out to be typically quite resistant to later attempts at correction

[86]

---

[87]

(the equivalent of our post-warnings). It seems unlikely of course that the post-warning procedures used in eyewitness settings can be directly translated into corrections of initial misinformation, but there are some interesting resemblances. For instance, Lewandowsky et al. (2012, p. 117) state that "the continued influence of misinformation can be eliminated through the provision of an alternative account that explains *why* the information was incorrect", which reminds of the logic of the enlightenment procedure that also gives reasons for the presence of misinformation. Looking at procedures that have been used in the other misinformation field might inspire researchers in both fields and help develop (even more) effective debiasing procedures.

---

[8] A particularly ironic example of this – in the present context – is contained in the Wikipedia entry for the (eyewitness) misinformation effect (http://en.wikipedia.org/wiki/Misinformation_effect; retrieved 17 August 2013): "If participants are warned prior to the presentation of misinformation, they are often able to resist misinformation's influence. However, if warnings are given after the presentation of misinformation, they do not aid participants in discriminating between original and post-event information."

References

[Studies included in the meta-analysis are marked with an *.]

Bekerian, D. A., & Bowers, J. M. (1983). Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 139-145.

Belli, R. F. (1989). Influences of misleading postevent information: Misinformation interference and acceptance. *Journal of Experimental Psychology: General, 118,* 72-85.

* 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.

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.

* Blank, H. (1998). Memory states and memory tasks: An integrative framework for eyewitness memory and suggestibility. *Memory, 6,* 481-529.

Blank, H. (2005). Another look at retroactive and proactive interference: A quantitative analysis of conversion processes. *Memory, 13,* 200-224.

Blank, H. (2009). Remembering: A theoretical interface between memory and social psychology. *Social Psychology, 40,* 164-175.

Blank, H., Ost, J., Davies, J., Jones, G., Lambert, K., & Salmon, K. (2013). Comparing the influence of directly vs. indirectly encountered post-event misinformation on eyewitness remembering. *Acta Psychologica, 144*, 635-641.

* Bodner, G. E., Musch, E., & Azad, T. (2009). Reevaluating the potency of the memory conformity effect. *Memory & Cognition, 37,* 1069-1076.

Bowers, J. M., & Bekerian, D. A. (1984). When will postevent information distort eyewitness testimony? *Journal of Applied Psychology, 69,* 466-472.

Bregman, N. J. & McAllister, H. A. (1982). Eyewitness testimony: The role of commitment in increasing reliability. *Social Psychology Quarterly, 45,* 181-184.

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.

Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning & Verbal Behavior, 19,* 695-704.

* 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., Groll, S. & Hirst, W. (2007). Tainted truth: Overcorrection for misinformation influence on eyewitness memory. *Social Cognition, 25,* 367-409.

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

* 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. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17,* 533-543.

Geiselman, R. E., Fisher, R. P., Cohen, G., Holland, H., & Surtes, L. (1986). Eyewitness responses to leading and misleading questions under the cognitive interview. *Journal of Police Science and Administration, 14,* 31-39.

* Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning & Verbal Behavior, 21,* 207-219.

Haddock, C. K., Rindskopf, D. & Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3,* 339-353.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego: Academic Press.

* Higham, P. A. (1998). Believing details known to have been suggested. *British Journal of Psychology, 89,* 265-283.

* Higham, P. A., Luna, K., & Bloomfield, J. (2011). Trace-strength and source-monitoring accounts of accuracy and metacognitive resolution in the misinformation paradigm. *Applied Cognitive Psychology, 25,* 324-335.

* Highhouse, S., & Bottrill, K. V. (1995). The influence of social (mis)information on memory for behavior in an employment interview. *Organizational Behavior and Human Decision Processes, 62,* 220-229.

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

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

Lampinen, J. M. & Smith, V. L. (1995). The incredible (and sometimes incredulous) child witness: Child eyewitnesses' sensitivity to source credibility cues. *Journal of Applied Psychology, 80,* 621-627.

Lane, S. M., Roussel, C. C., Villa, D., & Morita, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source-monitoring errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1131-1142.

Lewandowski, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13,* 106-131.

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

Lindsay, D. S. (2008). Source monitoring. In H. L. Roediger, III (Ed.), *Cognitive psychology of memory* (pp. 325-348). Oxford: Elsevier.

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

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

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

Loftus, E. F. (1991). Made in memory: Distortions of recollection after misleading information. In G. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 187-215). New York: Academic Press.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning and 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.

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.

* Meade, M. L. & Roediger, H. L. III. (2002). Explorations in the social contagion of memory. *Memory & Cognition, 30,* 995-1009.

* Oeberst, A. & Blank, H. (2012). Undoing suggestive influence on memory: The reversibility of the eyewitness misinformation effect. *Cognition, 125,* 141-159.

Pansky, A., Tenenboim, E., & Bar, S. K. (2011). The misinformation effect revisited: Interactions between spontaneous memory processes and misleading suggestions. *Journal of Memory and Language, 64,* 270-287.

* Paterson, H. M., Kemp, R., & McIntyre, S. (2012): Can a witness report hearsay evidence unintentionally? The effects of discussion on eyewitness memory. *Psychology, Crime & Law, 18,* 505-527.

* Paterson, H. M., Kemp, R. I., & Ng, J. R. (2011). Combating co-witness contamination: Attempting to decrease the negative effects of discussion on eyewitness memory. *Applied Cognitive Psychology, 25,* 43-52.

Payne, D. G., Toglia, M. P., & Anastasi, J. S. (1994). Recognition performance level and the magnitude of the misinformation effect in eyewitness memory. *Psychonomic Bulletin & Review, 1,* 376-382.

Poole, D., & Lindsay, D. S. (2002). Reducing child witnesses' false reports of misinformation from parents. *Journal of Experimental Child Psychology, 81,* 117-140.

Popper, K. (1934). *Logik der Forschung*. Vienna: Julius Springer. [English translation (1959): *The logic of scientific discovery*. London: Hutchinson.]

* Price, H. L. & Connolly, D. A. (2004). Event frequency and children's suggestibility: A study of cued recall responses. *Applied Cognitive Psychology, 18,* 809-821.

* Principe, G. F., Haines, B., Adkins, A., & Guiliano, S. (2010). False rumors and true belief: Memory processes underlying children's errant reports of rumored events. *Journal of Experimental Child Psychology, 107,* 407-422.

Smith, V. L. & Ellsworth, P. C. (1987). The social psychology of eyewitness accuracy: Misleading questions and communicator expertise. *Journal of Applied Psychology, 72,* 294-300.

Sporer, S. L. (1982). A brief history of the psychology of testimony. *Current Psychological Reviews, 2,* 323-339.

* Szpitalak, M. & Polczyk, R. (2010). Warning against warnings: Alerted subjects may perform worse. Misinformation, involvement and warning as determinants of witness testimony. *Polish Psychological Bulletin, 41,* 105-112.

* Szpitalak, M. & Polczyk, R. (2011). Can warning harm memory? The impact of warning on eyewitness testimony. *Problems of Forensic Sciences, 86,* 140-150.

* Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language, 63,* 149-157.

Tulving, E. (1983). *Elements of episodic memory.* Clarendon Press: Oxford.

* Wright, D. B. (1993). Misinformation and warnings in eyewitness testimony: A new testing procedure to differentiate explanations. *Memory, 1,* 153-166.

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: Exploring misinformation effects when presented by another person. *British Journal of Psychology, 91,* 189-202.

Zaragoza, M. S., Belli, R. S., & Payment, K. E. (2006). Misinformation effects and the suggestibility of eyewitness memory. In M. Garry & H. Hayne (Eds.), *Do justice and let the sky fall: Elizabeth F. Loftus and her contributions to science, law, and academic freedom* (pp. 35-63). Hillsdale, NJ: Lawrence Erlbaum Associates.

* 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.

Appendix: *Included Studies, General Study Characteristics, Warning Dimensions, Memory Performances and Log Odds Ratios*

Study/ Experiment/ Condition	No. Participants (1)	Type of Participants (2)	Misinformation Paradigm (3)	Delay of Misinformation (4)	No. of Misleading Details (5)	Type of Misleading Details (6)	Delay of Test (4)	Type of Test (7)	Warning Specificity (8)	Enlightenment	Social Discrediting	Memory for Original Event Details (% correct)				Log Odds Ratio (9)		Misinformation Endorsement (% endorsed)				Log Odds Ratio (9)	
												Misled		Control				Misled		Control			
												Post-warning	No warning	Post-warning	No Warning	Post-warning	No Warning	Post-warning	No Warning	Post-warning	No warning	Post-warning	No Warning
Be94-Ex1	72/--	STU	STA	Short	2	CON	Short	REC	PRES	No	No	37	--	47	--	0.43	--						
Be94-Ex2	77/--	STU	STA	Short	2	CON	Short	REC	PRES	No	No	10	--	16	--	0.54	--	21	--	6	--	1.43	--
Be94-Ex3	288/--	STU	STA	Short	2	CON	Short	REC	PRES	No	No	26	--	36	--	0.45	--						
Be94-Ex4	40/--	STU	STA	Short	3	CON	Short	REC	PRES	No	No	27	--	43	--	0.71	--						
Bl98-Ex1	22/24	STU	STA	Medium	4	CON	Short	FC	PRES	Yes	No	78	50	78	78	0.00	1.27	22	50	22	22	0.00	1.27
Bl98-Ex2	73/--	STU	STA	Medium	4	CON	Short	MST	PRES	Yes	No	57	--	60	--	0.11	--	6	--	9	--	-0.33	--
Bo09 dyads	49/55	STU	DP	Short	~1.5	SUP	Short	REC	POSS	No	No							29	66	10		1.30	2.86
Bo09 written report	49/59	STU	STA	Short	~1.5	SUP	Short	REC	POSS	No	No							21	49			0.87	2.16
Ch83-Ex1	34/31	STU	STA	Long	4	CON	Medium	FC	PRES	No	No	95	41	95		0.00	3.31	5	59	5		0.00	3.31
Ch83-Ex2	40/43	STU	STA	Long	4	CON	Medium	FC	PRES	No	No	43	21	36		-0.29	0.75	47	74	32		0.63	1.80
Ea03-Ex1 MRT	48/--	STU	STA	Medium	2	CON	Medium	MOD	LOPP	No	No	69	--	68	--	-0.05	--						
Ea03-Ex1 MOT	48/--	STU	STA	Medium	2	CON	Medium	REC	IDEN	No	No	44	--	60	--	0.65	--						
Ea03-Ex2	80/--	STU	STA	Medium	4	CON	Medium	REC	IDEN	No	No	40	--	68	--	1.16	--						
Ea03-Ex3	72/--	STU	STA	Medium	4	CON	Medium	REC	IDEN	No	No	46	--	63	--	0.69	--						

Ea03-Ex4A low	~114/--	STU	STA	Medium	4	CON	Medium	REC	IDEN	No	No	37	--	44	--	0.29	--						
Ea03-Ex4A high	~114/--	STU	STA	Medium	4	CON	Medium	REC	IDEN	No	No	37	--	49	--	0.49	--						
Ea03-Ex4B low	~114/--	STU	STA	Medium	4	CON	Medium	REC	IDEN	No	No	55	--	71	--	0.69	--						
Ea03-Ex4B high	~114/--	STU	STA	Medium	4	CON	Medium	REC	IDEN	No	No	54	--	71	--	0.74	--						
Ec05-Ex1 untrustworthy	29/34	STU	STA	Short	4	SUP	Short	REC	POSS	No	Yes							32	46	9	5	1.56	2.78
Ec05-Ex1 incompetent	28/34	STU	STA	Short	4	SUP	Short	REC	POSS	No	Yes							24		13		0.75	
Ec05-Ex2	20/20	STU	STA	Short	8	SUP	Short	YN	POSS	No	Yes							13	26	8	6	0.54	1.71
Ec05-Ex3 social	24/23	STU	STA	Short	4	SUP	Short	REC	POSS	No	Yes							24	41	5	5	1.79	2.58
Ec05-Ex3 explicit	23/23	STU	STA	Short	4	SUP	Short	REC	PRES	No	No							19		6		1.30	
Ec05-Ex4 social	30/29	STU	STA	Short	4	SUP	Short	YN	POSS	No	Yes							26	42	22	22	0.22	0.94
Ec05-Ex4 explicit	29/29	STU	STA	Short	4	SUP	Short	YN	PRES	No	No							23		19		0.24	
Ec07-Ex1 social	20/21	STU	STA	Short	16	SUP	Short	YN	PRES	No	Yes							37	49	26	20	0.51	1.35
Ec07-Ex1 explicit	18/21	STU	STA	Short	16	SUP	Short	YN	PRES	No	No							39		25		0.65	
Ec07-Ex2 social	30/28	STU	STA	Short	16	SUP	Short	YN	PRES	No	Yes							29	50	21	25	0.43	1.10
Fr02-Ex1&2 10 min	24/24	STU	STA	Short	2	CON	Short	YN	PRES	No	No	45	46	64	70	0.78	1.01	25	30	3	3	2.56	2.82
Fr02-Ex1&2 1 week	24/24	STU	STA	Short	2	CON	1 week	YN	PRES	No	No	28	33	55	58	1.15	1.03	35	38	3	3	3.04	3.17
Gr82-Ex1	~18/~18	STU	STA	Short	4	CON	Short	FC	POSS	No	Yes	28	22	58	90	1.27	3.49						
Hi98-Ex1 short	28/--	STU	STA	Long	5	SUP	Medium	SMT	PRES	No	No							34	--	23	--	0.55	--
Hi98-Ex1 long	28/--	STU	STA	Medium	5	SUP	Long	SMT	PRES	No	No							40	--	20	--	0.98	--
Hi98-Ex2	46/--	STU	STA	Medium	5	SUP	Medium	SMT	PRES	No	No							39	--	27	--	0.55	--
Hi11-Ex2 low	20/--	STU	STA	Short	12	CON	Short	FC	PRES	No	No	64	--	78	--	0.67	--	36	--	23	--	0.67	--
Hi11-Ex2 high	20/--	STU	STA	Short	12	CON	Short	FC	PRES	No	No	63	--	77	--	0.70	--	38	--	23	--	0.70	--

Hi95	59/70	STU	STA	Short	10	CON	Short	YN	PRES	Yes	No							58	83	54		0.16	1.43
Li90 low	68/--	STU	STA	Short	3	CON	Long	REC	LOPP	No	No	45	--	51	--	0.24	--	27	--	9	--	1.32	--
Li90 high	68/--	STU	STA	Long	3	CON	Short	REC	LOPP	No	No	39	--	48	--	0.37	--	13	--	10	--	0.30	--
Li95-Ex2 pre low	24/24	CHI	STA	Long	2	CON	Long	FC	LOPP	No	No	90	79	88	94	-0.21	1.38	10	19	6	3	0.55	1.98
Li95-Ex2 pre high	24/24	CHI	STA	Long	2	CON	Short	FC	LOPP	No	No	75	58	92	77	1.34	0.88	15	35	4	12	1.36	1.44
Li95-Ex2 third low	~18/~18	CHI	STA	Long	2	CON	Long	FC	LOPP	No	No	94	97	94	97	0.00	0.00	3	7	3	3	0.00	0.71
Li95-Ex2, third high	~18/~18	CHI	STA	Long	2	CON	Short	FC	LOPP	No	No	81	70	97	93	2.12	1.67	19	30	1	4	2.83	2.41
Me02-Ex1	18/18	STU	DP	Short	6	SUP	Short	REC	POSS	No	No							17	36	6	7	1.22	2.09
Oe12-Ex1 (a)	55/--	STU	STA	Short	2	both	Short	MST	PRES	Yes	No	77	--	74	--	-0.16	--						
Oe12-Ex1 (b)	63/--	STU	STA	Short	2	both	Short	MST	PRES	Yes	No	72	--	70	--	-0.10	--						
Oe12-Ex2	26/26	STU	STA	Short	2	both	Short	FC	PRES	Yes	No	85	63	83	81	-0.15	0.92	15	37	17	19	-0.15	0.92
Oe12-Ex3 (a)	28/--	STU	STA	Short	4	both	5 weeks	FC	PRES	Yes	No	71	--	83	--	0.69	--	29	--	17	--	0.69	--
Oe12-Ex3 (c)	26/--	STU	STA	Short	4	both	5 weeks	MST	PRES	Yes	No	75	--	63	--	-0.57	--						
Pa11-Ex1	32/32	STU	DP	Short	~5	both	1 week	REC	POSS	No	No							10	9	4	4	1.03	0.97
Pa11-Ex2	24/24	STU	DP	Short	~5	both	1 week	REC	POSS	No	No							28	22	3	3	2.42	2.06
Pa12-Ex1 specific	34/34	STU	DP	Medium	~4	SUP	1 week	REC	POSS	No	Yes							11	9	1	1	2.79	2.54
Pa12-Ex2	32/32	STU	DP	Medium	~4	SUP	1 week	REC	POSS	No	Yes							9	11	1	1	2.60	2.82
Pr04 single	~15/~15	CHI	DP	Long	6	CON	Long	REC	LOPP	No	Yes	49	50	47	48	-0.09	-0.09	5	1	2	2	0.84	-0.55
Pr04 repeated	~15/~15	CHI	DP	Long	6	CON	Long	REC	LOPP	No	Yes	21	11	21	21	0.00	0.76	2	3	2	2	0.08	0.52
Pr10 overheard 3-4 y	44/45	CHI	STA	Medium	1	SUP	1 week	REC	LOPP	No	No							65	92			4.37	6.14
Pr10 classmate 3-4 y	44/42	CHI	DP	Medium/ Long	1	SUP	Long/ 1 week	REC	LOPP	No	No							96	90	2		6.83	5.99
Pr10 overheard 5-6 y	42/42	CHI	STA	Medium	1	SUP	1 week	REC	LOPP	No	No							55	95	2		3.94	6.68

Pr10 classmate 5-6 y	44/42	CHI	DP	Medium/ Long	1	SUP	Long/ 1 week	REC	LOPP	No	No							64	95			4.30	6.68
Sz10-Ex1 low	96/87	STU	STA	Short	13	both	Short	YN	PRES	No	No							63	63	47	53	0.66	0.41
Sz10-Ex1 high	73/93	STU	STA	Short	13	both	Short	YN	PRES	No	No							44	69	53	41	-0.37	1.13
Sz11	95/96	STU	STA	Short	8	CON	Short	FC	PRES	No	No	58	48	63	77	0.18	1.26	42	52	37	23	0.18	1.26
Th10-Ex1 single	20/20	STU	STA	Medium	8	CON	Short	REC	POSS	No	No	58	44	66	67	0.32	0.93	19	30	5	2	1.41	2.92
Th10-Ex1 repeated	20/20	STU	STA	Medium	8	CON	Short	REC	POSS	No	No	69	28	75	72	0.27	1.86	20	48	3	5	2.02	2.77
Th10-Ex2 single	18/17	STU	STA	Medium	8	CON	Short	FC	POSS	No	No	59	62	79	82	0.93	0.99	31	31	9	11	1.49	1.34
Th10-Ex2 repeated	14/17	STU	STA	Medium	8	CON	Short	FC	POSS	No	No	67	41	80	82	0.68	1.88	24	53	10	9	1.00	2.47
Wr93	~102/~102	STU	STA	Short	1	CON	Short	FC	IDEN	No	No	81	51	86	86	0.37	1.78	2	43	4	4	-0.71	2.90
Zaragoza & Lane (1994)	~66/~66	STU	STA	Short	5	SUP	Short	SMT	PRES	No	No							35	30	15	16	1.12	0.81

Notes. (1) *N*(post-warning condition)/*N*(no-warning condition). (2) STU = students, CHI = children. (3) STA = standard misinformation paradigm using written (or sometimes audiotaped) misinformation, DP = direct personal introduction of misinformation (cf. Table 1 and footnote 2). (4) Short = <20 min, Medium = 20 min to 1 h, Long = 1 to 4 days. (5) No. of misleading details encountered by any one participant (not necessarily identical with the total number of critical items). (6) CON = contradictory, SUP = supplementary, both = both types. (7) REC = recall; all other tests are some form of recognition test: FC = (standard) forced-choice recognition (between 2-4 alternatives, and including the MI as a response alternative), MOD = modified recognition (forced-choice, *excluding* the MI as a response alternative), MST = memory state test (see Blank, 1998), SMT = source monitoring test, YN = yes/no test. (8) POSS = participants alerted to possible presence of MI, PRES = presence of MI positively assured, LOPP = logic of opposition-type warning, IDEN = explicit identification of the MI. (9) Because all calculations were based on them, we give the log odds ratios here (instead of the odds ratios). Shaded cells highlight cases where two-alternative forced-choice recognition tests had been used and therefore the effects can be coded as both Original Memory and Misinformation Endorsement.

Notes on individual studies.

Be94 = Belli, Lindsay, Gales & McCarthy (1994). In all of Belli et al.'s (1994) experiments, memory for both original and misleading details was assessed; only in Exp. 2, however, this was combined with source attribution, thus permitting to speak (in the case of misattribution to the original event source) of Misinformation *Endorsement*. Exp. 3 had two separate control conditions using neutral and no information, respectively; the control performance given here is averaged across these. In Exp. 4, another 24 participants were not shown critical original details and therefore do not qualify for our analysis. The memory performances in Exp. 4 are estimated from Figure 2 in Belli et al. (1994).

Bl98 = Blank (1998). The memory performances in Exp. 2 are aggregated across several experimental and control conditions (that used different sources of original and misleading information; cf. Fig. 4 in Blank, 1998). The Misinformation Endorsement percentages, specifically, are percentages of misattributions of misleading post-event details to the source of the original information (taken from Fig. 4 in Blank, 1998).

Bo09 = Bodner, Musch & Azad (2009). Bodner et al.'s Exp. 1 involved no-warning and control groups, and Exp. 2 involved post-warning groups only; as the two experiments were identical in all other respects, we used the Exp. 1 control groups as controls for the Exp. 2 post-warning groups. The number of encountered misleading details depended on what was

actually reported (in the co-witness discussion or in the written witness report) and differed slightly across groups (means ranging from 1.48 to 1.69); the reported number is a rough average. The misled memory performances are based on the authors' 'misinformation index' (not on the information reported in their Table 1). The control performances were calculated on the basis of the average number of presented details in the other conditions (see above), as well as on additional raw data provided by Glen Bodner.

Ch83 = Christiaansen & Ochalek (1983).

Ea03 = Eakin, Schreiber & Sergent-Marshall (2003). Exp. 1: MRT = modified recognition test (Eakin et al.'s acronym; same as MOD in this table, see note 7); MOT = modified opposition test (Eakin et al.'s acronym). Exps. 3 & 4: These comparisons contain only the *control* and the *misled-warning at test only* conditions. The memory performances are estimated from Figures 2, 4, and 6 in Eakin et al. (2003). Exp. 4: 'low'/'high' means low and high accessibility of the misinformation. The *N*s are estimated from the total *N*s in Exps. 4A and 4B.

Ec05 = Echterhoff, Hirst & Hussy (2005). Exp. 1: 'untrustworthy' and 'incompetent' refer to specific ways of discrediting the misinformation source. Exp. 2: This comparison contains only the *control* and *social postwarning* conditions. Exps. 3 & 4: 'social' and 'explicit' is short for 'social postwarning' and 'explicit monitoring' (the authors' labels).

Ec07 = Echterhoff, Groll & Hirst (2007). Exps. 1 & 2: 'social' and 'explicit' – see Ec05. Exp. 2: The *source monitoring task* condition is not included in this comparison.

Fr02 = Frost, Ingraham & Wilson (2002). Exps. 1 and 2 were identical except for the absence (Exp. 1) or presence (Exp. 2) of a warning; hence they were combined for this comparison. As one of the control groups in Frost et al. (2002) had a value of exactly 0%, it was impossible to calculate effect sizes for this cell; therefore, we decided to pool the estimates of all four respective cells (which were estimates of the exact same thing, namely, the tendency to guess misleading details when they had not in fact been presented; the original values ranged from 0-4%).

Gr82 = Greene, Flynn & Loftus (1982). Only Exp.1 contained a post-warning condition, all other experiments in Greene et al. (1982) used pre-warnings. *N*s are estimated from total *N* in Exp. 1; two other conditions using pre-warnings only are not included here. Memory performances are estimated from Figure 3 in Greene et al. (1982).

Hi98 = Higham (1998). Exp. 1: 'short' and 'long' refer to short (medium in our classification) and long delays between misinformation presentation and test.

Hi11 = Higham, Luna & Bloomfield (2011). Exp. 1 used a combination of a pre- and a post-warning and is therefore not included here. Exp. 2: 'low' and 'high' refer to low and high incentives for participants to provide accurate memory test answers. The memory performances are collapsed across fine- and coarse-grained answers as well as across testify option (as none of these distinctions figured in other studies; see Higham et al., 2011, for details; collapsed data provided by Phil Higham).

Hi95 = Highhouse & Bottrill (1995).

Li90 = Lindsay (1990). 'Low' and 'high' refer to low and high discriminability of the sources of original event information and post-event misinformation.

Li95 = Lindsay, Gonzales & Eso (1995). Exp. 1 used a within-participants manipulation of warning and was therefore not included here. Exp. 2: 'Pre' and 'third' refer to preschoolers and third graders; 'low' and 'high' refer to low and high discriminability of the sources of original event information and post-event misinformation. The precise *n*s for the third graders are specified in the paper as a range only (i.e. 15-20); we used an average for our analyses.

Me02 = Meade & Roediger (2002). Only Exp. 1 in Meade and Roediger (2002) contained a post-warning condition. All recall percentages are averaged across the low and high expectancy items listed separately in their Table 1; as this distinction features in no other study considered here, it did not make sense to list these results separately; similarly, their distinction between *remember* and *know* answers is ignored for the present purposes.

Oe12 = Oeberst & Blank (2012). Exp. 1: (a) and (b) refer to the *ignorant-enlightened* and *enlightenment only* groups. Exp. 2: Only the *ignorant-throughout* and *enlightened only* groups were used in this warning/no warning comparison; as the *ignorant-enlightened* group was first tested under standard conditions and then post-warned and tested a second time, their performance (in the respective tests) could have been arbitrarily added to either the warning or no-warning conditions; therefore we decided to exclude it altogether. Exp. 3: (a) and (c) refer to the *ignorant-enlightened* and *late enlightenment* groups.

Pa11 = Paterson, Kemp & Ng (2011). Exps. 1 & 2: Each co-witness in a dyad watched a slightly different videotape, differing from the other version in 4 contradictory details and 4 or 2 (depending on videotape version) supplemental details; hence, every participant may have encountered 7 pieces of misinformation on average; however, it is not reported how many of

these actually transpired in each co-witness discussion; on the basis of the ratios reported in Bodner et al. (2009; see note above) we estimate the number of encountered misleading detail in this study to be approximately 5. Exp. 2: No-warning and delayed warning groups only.

Pa12 = Paterson, Kemp & McIntyre (2012). Exp. 1: 'Specific' refers to the 'specific warning' condition. The 'general warning' condition was not used, as this procedure did not meet our definition of a warning. Also, the 'no discussion' control condition was disregarded. Exps. 1 & 2: We used the 'same video' conditions as controls for the 'different video' (i.e., misled) conditions. We estimated the number of misleading that actually transpired in the discussion to be around 4, following the same rationale as for Pa11. The memory performances then resulted from dividing the average absolute numbers of endorsed misleading details (estimated from their Figures 2 and 6) by four. Further, because of empty cells, the control performances were pooled across both Exps. 1 and 2 and across the post-warning and no-warning conditions (see Fr02 for the same problem and solution).

Pr04 = Price & Connolly (2004). 'Single' and 'repeated' refer to single or repeated original events experienced by the children. All cell *N*s are inferred from the total *N*; children in the *moderate instructions* conditions are not included in our analysis; although these conditions did not use warnings, they did not correspond to a standard no-warning condition either. Both free and cued recall data are reported in the paper; we use only the former because this was the first memory test conducted. All Misinformation Endorsement free recall control performances were originally at a level of 0%, which would make it impossible to calculate our effect sizes; we therefore set all these performances to an approximate average level obtained in other studies with similar problems (2%; derived from Fr02, Pa12 and Pr10).

Pr10 = Principe, Haines, Adkins & Guiliano (2010). 'Overheard' and 'classmate' refer to different ways of encountering misinformation; '3-4 y' and '5-6 y' denotes the age of the investigated children. As the 'classmate' condition involved encountering the misinformation as a rumour from classmates at some point between planting of the misinformation and the final memory test, no clear delays could be determined. Control Misinformation Endorsement of the 3-4 year olds was at 0%, making it impossible to calculate our effect sizes; therefore we estimated the control performance by including the performance of the 5-6 year olds (originally 5%; see Fr02 for the same problem and solution).

Sz10 = Szpitalak & Polczyk (2010). 'Low' and 'high' refer to low or high involvement of participants in the topic of the original event. The results reported in the original paper reflect a mix of Original Memory (3 critical items) and Misinformation Endorsement (10 critical items - 4 contradictory and 6 supplemental); we take only Misinformation Endorsement into account (based on original data provided by Romuald Polczyk).

Sz11 = Szpitalak & Polczyk (2011).

Th10 = Thomas, Bulevich & Chan (2010). 'Single' and 'repeated' refer to testing; i.e. participants receiving no or a prior memory test (without warning) before the final (post-warned) memory test. The cell *N*s, as well as the Misinformation Endorsement control performances, were provided by Ayanna Thomas. Two types of control conditions were used throughout, one providing no post-event information on critical details and one providing consistent information; as these conditions were within-participants and of minor interest for us, we averaged performances across them.

Wr93 = Wright (1993). The cell *N*s are estimated from the total *N*. Misinformation Endorsement in both the misled-warned and control-neutral conditions was exactly 0%, making it impossible to calculate odds ratios; therefore, we replaced these values with 2% (see Pr04 for the same problem and solution). Control Misinformation Endorsement was then averaged across a control-repeated and a control-neutral condition (cf. Be94, Th10).

Za94 = Zaragoza & Lane (1994). The cell *N*s are estimated from the total *N*.