Remembering discrepant information in eyewitness testimony

experiments and paired-associate learning tasks:

an integrative model

Hartmut Blank

Eingereicht als Dissertation an der Sozialwissenschaftlichen Fakultät der Universität Konstanz zur Erlangung des akademischen Grades des Doktors der Sozialwissenschaften

---

Referenten:

Prof. Dr. Peter Steck, Universität Konstanz

Prof. Dr. Wolfgang Hell, Universität Münster

Die Dissertation erschien 1996 in Mikrofiche-Ausgabe als Band 11 der Reihe Psychologie in der "Edition Wissenschaft" des Tectum-Verlags, Marburg. ISBN 3-89608-581-6

Part of the research reported in this dissertation was supported by the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung* (Projekt-Nr. P8842-MED). Furthermore, I benefitted from a 27-month *Promotionsstipendium* of the state of Baden-Württemberg. Experiment 1 in part I (experiment 3 in part I / experiment 1 in part II) was presented in 1991 (1993 / 1994) at the 33. (35. / 36.) *Tagung experimentell arbeitender Psychologen* in Gießen (Münster / Munich), Germany. I am grateful to Margarete Boos, Wolfgang Hell, Ralph Hertwig, Ulrich Hoffrage, Martin Spieß, and Peter Steck for helpful comments on various parts of this work and for occasional encouragement; Willi Nagl provided statistical advice. Also, I would like to thank the Tourist Information, Constance, as well as Georg Baur and Christoph Danneffel, for their help in the preparation of slides and Axel Renner for computer advice.

Content

Summary

Part I of this dissertation is concerned with the eyewitness misinformation effect: When subjects are given misleading post-event information after having witnessed a simulated event, they tend to report the suggested details when tested for their memory of the event. An integrative model is presented which depicts performance in such eyewitness suggestibility experiments as the subjects' solutions of memory tasks, depending on (a) a specified task-relevant memory base and (b) subjects' perception of the memory task. Three existing accounts of the effect of misleading post-event information are reinterpreted in terms of the model and reduced to one single core: subjects answer test questions while assuming the consistency of event and post-event information. Employing methodological innovations suggested by the model, experiments 1 and 2 examined the memory base and found no evidence for memory impairment. In Experiment 3, the usual misinformation effect obtained with the standard test procedure of Loftus, Miller, and Burns (1978) disappeared when destroying the subjects' consistency assumption prior to testing. A qualitative analysis of subjects' discrepancy detections shows that suggestibility (resistance) to misleading information was associated with a (non-) consistency assumption. A follow-up study of experiments 2 and 3 after months revealed effects of misleading information which might reflect long-time integration of information due to source loss. The general discussion of part I relates these findings to eyewitness suggestibility research in general, focusing on memory impairment, the transformation of memory representations into memory reports, and the relation of memory integration to source memory and discrepancy detection.

In part II, the model is adapted to paired-associate learning. Based on model simulations, test-dependent retroactive and proactive inhibition effects on recognition memory are predicted, given that (a) there is poor memory for the list membership of responses and (b) a local test procedure is used that asks for responses from only one of the lists at a time. Conversely, with a global test procedure that simultaneously asks for both responses associated with a stimulus term, no such effects are predicted. The predictions were confirmed in two experiments, using a two-alternative forced-choice test (experiment 1) or a multiple choice test with a "don't know" option (experiment 2) as the local test procedure. The results were further analyzed in terms of the model and revealed a reasonably good fit of post-hoc simulations to the actual data. Implications of these findings for research on paired-associate learning and for eyewitness interrogation procedures are discussed.

Introduction

The work presented in this doctoral dissertation has undergone some evolutionary changes in its progress. I started with an interest in the classic question of modern research on eyewitness memory, namely, the question whether memory for details of a witnessed event can be impaired by subsequent discrepant information. This interest is reflected most clearly in the experiments 1 and 2 in part I of this work (actually, experiment 1 was my *Diplomarbeit*; it is reported here because it is a natural starting point for the following experiments and because some features of experiment 2 cannot be understood without reference to experiment 1). Accordingly, the integrative model presented in part I was initially motivated by the wish to find an approach to the measurement of potential memory impairment which circumvents certain problems of existing test procedures. To this end, it was necessary to theoretically seperate memory processes from processes that translate memory representations into performance in a memory test. Later, when I had arrived at the conclusion that memory impairment plays but a minor role with respect to eyewitness suggestibility, my focus gradually shifted to these translation processes in their own right. First, I became interested in deteriorations of memory performance which - according to the model - should result from the subjects' assumptions regarding the consistency of information presented to them in the experiment. As a matter of fact, most investigators in eyewitness suggestibility experiments deceive their subjects on the presence of misleading post-event information. Consequently, an investigation of this issue might help understanding certain aspects and processes in eyewitness suggestibility experiments - like response biases or the impact of discrepancy detection on test performance (see experiment 3 of part I). While this first investigation into non-impairment mechanisms of suggestibility yielded some striking results and insights, I was nevertheless left with the sad aftertaste that I had merely dealt with artifacts, that is, phenomena that are restricted to a special way of conducting laboratory research. Therefore, the next step was to look for effects of post-event information which were not tied to the deception issue. Relying on tentative evidence reported in part I and on model simulations, I speculated on misinformation effects which may result from an interaction of poor memory for the sources of presented details and certain aspects of the test procedure, even when there is no reason for subjects to assume the consistency of information. Though these evidence and simulations came from a laboratory context, it seems quite plausible that they should also apply to real eyewitness testimony, since forgetting of the sources of information is a pervasive everyday memory phenomenon. The investigation of such effects constitutes part II of this work. Because the simulations indicated that the effects should be small in magnitude and because of the explorative character of these studies, I decided to minimize any additional procedural variance inherent in more real-world-like designs (think of modality differences in the presentation of information in eyewitness designs, for instance) and to perform simple paired-associate learning experiments.

Note that there is a close resemblance between eyewitness misinformation designs and retroactive interference designs, leading some researchers to even speak of the misinformation effect as a special kind of retroactive inhibition (e.g., Chandler, 1989). Thus, the shift to paired-associate learning constitutes no major departure from the topic under consideration, and I do not doubt that the results of part II bear on real eyewitness testimony as well.

Before presenting my research, I will introduce the reader into the fields of eyewitness suggestibility and paired-associate learning. This will be done in more detail for the former, whereas the literature on the latter topic is abundant, so that only some remarks will be necessary.

The eyewitness suggestibility "research landscape"

I am concerned here with modern research on eyewitness suggestibility and eyewitness memory, which has enormously expanded since the mid-seventies, stimulated largely by the work of Elizabeth Loftus and her coworkers. Although there has been extensive early work on this topic, particularly in Germany around the turn of the century, an account of this research is beyond the scope of a brief introduction (see Sporer, 1982, for a review). Further, I will not deal here with the various theoretical approaches to eyewitness suggestibility and eyewitness memory, since these are worked through at length in part I of this work. Rather, I will try to give the reader a very rough first impression of the "research landscape" as it has developed until the present time, using some more or less classic examples from the literature and focusing somewhat on issues which are of interest for this work. This research landscape, in my view, has evolved along two main lines (which may be called the *pure suggestibility* line and the *memory* line), corresponding to two major sources of eyewitness suggestibility that have been of interest to researchers: Eyewitness reports may be distorted (1) through various aspects of the questioning process itself and (2) by post-event information, that is, additional information to which the witness is exposed in the time between the witnessed event and the testimony. Though the distinction between these two lines of research is seldom recognized, I consider it useful for an understanding of the development of the field and the phenomena under investigation.

As a starting point, let us take a look at two prototypical research designs. The first may be illustrated through a study by Loftus and Palmer (1974): Subjects were shown a filmed automobile accident and afterwards asked some questions about it. One of them required the subjects to estimate the speed of the cars at the moment of their collision. The critical manipulation was the wording of this question. For some subjects, it read: "About how fast were the cars going when they smashed into each other?" For other subjects, the verb *smashed* was replaced with either *collided*, *bumped*, *hit*, or *contacted*. It turned out that the subjects' speed estimates closely mirrored the severity of the verb used in the question: *smashed* yielded a mean speed estimate of 40,8 miles per hour, whereas *contacted* produced an estimate of 31,8 miles; the other verbs yielded values in between. For our purposes, it is important to see that the manipulation intended to influence the subjects' reports takes place at the very time the testimony ist given, that is, during the test phase. This constitutes a major difference to a second type of design which involves a three-phase procedure. In the first phase, just like above, the subjects witness an event, and in the last phase they take a memory test. Interpolated, however, is an additional phase where the critical manipulation takes place. The subjects are now given additional, post-event information which supplements or directly contradicts the witnessed event, and the question is whether this manipulation affects the subjects' reports at a later point in time, that is, in the test phase. A critical feature of this type of design is that the effect of the manipulation is delayed and, therefore, must exert its influence through (long-term) memory (how this is accomplished is a matter of heated debate, of course, but for the present purpose it is only necessary to realize that memory *is* involved).

Thus, following the terminology proposed above, we may speak of *pure suggestibility designs* consisting of two main phases and where the effect of the manipulation is immediate and *memory designs* consisting of three phases where the effect is delayed. Research of the first type has been more common in the beginnings of modern eyewitness research (e.g., Lipton, 1977; Loftus, 1975; Loftus & Zanni, 1975); later examples are Kallio and Cutler (1987), who investigated the impact of questions containing neutral or "unmarked" (e.g., tall) vs. biased or "marked" (e.g., short) adjectives on estimates of the heigth of a person, or Lipscomb, McAllister, and Bregman (1985) who assessed the influence of the same variable with respect to speed estimates. More loosely in this tradition, Moston (1990) gives a review of various sources of child suggestibility. This work has usually an applied touch and is (therefore?) typically found in lower-ranking journals. On the other hand, the mainstream of modern eyewitness suggestibility research as it has emerged in the 80's and which accounts for three quarters of the studies, at the very least, employs memory designs. One may speculate whether it was the very memory component that has made eyewitness suggestibility so attractive to researchers; it seems that it had offered the opportunity to re-fight old battles between different views of memory, particularly concerning the impact of interpolated information, on a new battleground, starting with Loftus's (e.g., 1979a) provoking claim that the performance decrements typically found after the introduction of misleading post-event information are due to an alteration of memory for the originally witnessed event, a claim that has not long gone unchallenged (see part I of this work). As a by-product of this linkage of memory theory with eyewitness suggestibility research, the new research field has found its way into higher-ranking, experimental journals and received a good amount of interest from the scientific community.

There is another feature of research designs which is correlated with the pure suggestibility vs. memory distinction. Pure suggestibility designs typically use continuous dependent variables

like estimates of personal attributes (age, height, weight, etc.) or speed estimates, and the suggestibility manipulation is usually indirect or unspecific in the sense that the information conveyed through the question (e.g., *smashed*) acts as a cue which permits inferences about a target variable (e.g., speed). Conversely, memory designs typically assess memory for discrete, easily identifiable items like traffic signs, tools, cigarette brands, etc., and the misleading post-event information is of the same kind (i.e., a traffic sign for another traffic sign). Of course, there are exceptions to this rule. For instance, Köhnken and Brockmann (1987) investigated the impact of unspecific post-event information (i.e., that a motorcyclist involved in an accident had been drinking or not) on a continuous variable (i.e., estimated speed of the motorcycle) with a memory design; further examples are Loftus (1977) and Belli (1988), who had their subjects estimate (on a virtually continuous scale) the color of certain objects after prior exposure to specific color misinformation. An exceptional pure suggestibility design is a study by Loftus and Zanni (1975) who investigated whether asking for a non-existing but plausible object using the definite vs. indefinite article ("Did you see the [vs. a] broken headlight?") affected subjects' choices on a yes/no recognition test with a "don't know" option. The same design (and also the same question) was used in a field study by Yuille and Cutshall (1986) which is especially worth mentioning since it is the only eyewitness misinformation study involving real eyewitnesses to crimes. Interestingly, these eyewitnesses (interviewed 4 or 5 months after the crime) were not susceptible to misinformation at all.

Though pure suggestibility designs play no major role in contemporary research on eyewitness memory (and none in this work), I find it nevertheless profitable to remind oneself of the simple fact showing up in these designs, namely, that subjects' responses can be influenced by suggestions, and without memory processes involved. In the face of the sheer mass of studies using a memory design, one can sometimes have the impression that the changes in people's recollections are solely a matter of memory processes, and McCloskey and Zaragoza's (1985) demonstration that suggestibility factors have their place in memory designs as well (see part I) would perhaps not have been such a shock had researchers not wholly shifted their attention to the memory component.

At the beginning of the "memory era", suggestibility was still a topic. Indeed, the introduction of the memory component was initially motivated by an interest for aftereffects of suggestibility manipulations. For example, one week after having influenced their subjects' speed estimates by suggesting that two cars had *smashed* into each other, Loftus and Palmer (1974) found that their subjects willingness to respond "yes" to the question: "Did you see any broken glass?" was also a function of the severity of the verb used in the initial question directed at the speed of the cars. Maybe the crucial step from suggestibility aftereffects towards a typical memory design, however, was the introduction of *specific* misleading details in questioning which could be probed directly in a later test, as Loftus (1975) did. One of the questions she asked her subjects after presentation of a short film clip ran: "Did you see

---

---

the children getting on the school bus?". In fact, there was no school bus in the film. After a week, 26% of these subjects responded "yes" to the question "Did you see a school bus?", compared to 6% of subjects who had not been asked the school bus question in the initial questionnaire. Thus, it appeared as if the suggested detail had been somehow integrated into subjects' memory of the film clip. A third group showed that suggestibility factors (I use this term rather loosely, meaning any manipulations that may change people's beliefs or underlying assumptions regarding an object or event in question) at least contributed to this result. Only 12% of subjects who had been asked directly for the school bus in the initial questionnaire ("Did you see a school bus?"), instead of having its existence presupposed in a question for something different, responded "yes" to the final critical question. However, instead of directly influencing the subjects' responses, as in previous studies, suggestibility factors were now assigned a different function: "... the subsequent information was introduced via presuppositions in questions, a technique which is effective in introducing information without calling attention to it" (Loftus, 1975, p. 571f). That is, the suggestibility manipulation was now considered to be only a means for effectively introducing misinformation into memory. Once introduced, memory processes were then held to operate and to be responsible for the subjects' responses on subsequent tests; potential carry-overs of the suggestibility manipulation to later test situations were neglected.

The last step to the full development of memory designs, finally, was an arrangement where suggested post-event details directly contradicted original event details and memory for these original (not the suggested) details was probed in the final test. This kind of design has seen enormous success; legions of experiments were carried out in this fashion. It shows a considerable similarity to retroactive interference designs (with the exception that there is no explicitly given stimulus term associated with the conflicting responses), and so it makes no wonder that some of the theoretical arguments developed in research on retroactive inhibition reappeared in new disguises (see part I). The classic example for this kind of research is a study by Loftus, Miller, and Burns (1978): Subjects were presented a series of slides depicting a car accident. One of the slides showed a car at an intersection with a stop sign. Thereafter, the subjects were asked some questions concerning the accident. The critical question introduced the misleading post-event detail: "Did another car pass the red Datsun when it was stopped at the yield sign?". 20 minutes later, the subjects were administered a two-alternative forced-choice recognition test where they were shown two versions of the critical slide, one containing the original stop sign and one showing the suggested yield sign. The result was straightforward: only 41% of the subjects chose the original slide when misled, compared to 75% correct choices of control subjects. This result has been replicated over and over in more than 20 published studies based on this design (Bekerian & Bowers, 1983; Bonto & Payne, 1991; Bowers & Bekerian, 1984; Bowman & Zaragoza, 1989; Ceci, Ross, & Toglia, 1987; Chandler, 1989; Dodson & Reisberg, 1991; Kroll, Ogawa, & Nieters,

1988; Loftus, 1991; Loftus, Donders, Hoffman, & Schooler, 1989; McCloskey & Zaragoza, 1985; Ober & Stillman, 1988; Pohl, Schumacher, & Friedrich, 1993; Shaughnessy & Mand; Sheehan & Grigg, 1985; Sheehan, Grigg, & McCann, 1984; Sheehan & Statham, 1989; Sheehan & Tilden, 1983, 1984, 1986; Tousignant, Hall, & Loftus, 1986; Trouvé & Libkuman, 1992; Wagenaar & Boer, 1987; Ward & Loftus, 1985; Weinberg, Wadsworth, & Baron, 1983; Zaragoza & Koshmider, 1989), even though there have also been some interesting failures to replicate it under certain conditions (see part I). That is, for some time, this procedure (which became known as the *standard test procedure*) was the method of choice for the investigation of eyewitness memory.

This changed drastically when McCloskey and Zaragoza (1985) demonstrated that the standard test procedure was far from being a proper measure for memory for original details but instead subject to response biases resulting from the suggestibility manipulation that introduced the misleading post-event information (see part I). These authors developed their own *modified test procedure*: Since the response biases are only possible when the suggested detail is a possible test response, the new procedure simply replaced it with a completely new alternative. McCloskey and Zaragoza argued that, if memory for original details is indeed impaired by subsequent misinformation, this would also lower performance on the modified test. They tested this prediction in a series of six experiments. Their subjects saw a slide series where a maintenance man enters an office, repairs a chair, and then steals some money and a calculator. There were four critical slides; one of them showed a coke can on a desk. Then, in a narrative, subjects read that a seven-up can had been on the desk, and the final test forced them to choose between coke and sunkist orange. With this new method, McCloskey and Zaragoza found no significant difference between experimental and control performance (72% vs. 75% correct across experiments). This null result has been replicated in a couple of studies (Belli, Windschitl, McCarthy, & Winfrey, 1992; Bonto & Payne, 1991; Bowman & Zaragoza, 1989; Ceci, Ross, & Toglia, 1987; Chandler, 1989, 1991; Loftus, 1991; Loftus, Donders, Hoffman, & Schooler, 1989), with occasional exceptions (see part I).

The standard test procedure and its successor, the modified test procedure, account for half of the research done with memory designs, at least. Predominant among the remaining half are yes/no-tests and recall procedures. Yes/no-tests are often used in studies that do not introduce discrepant but merely supplemental post-event information (like the study by Loftus, 1975, described above; see also Dodd & Bradshaw, 1980; Lassiter, Stone, & Rogers, 1988; Loftus, 1981; Read & Bruce, 1984; Schooler, Gerhard, & Loftus, 1986; Smith & Ellsworth, 1987), but there are also studies that assess memory for discrepant details with yes/no questions (Belli, 1989; Cole & Loftus, 1979; Loftus, Levidow, & Duensing, 1992; Pezdek & Greene, 1993; Tversky & Tuchin, 1989). Studies using (cued) recall procedures constitute a fairly heterogenous sample, typically with an applied touch, such as Geiselman, Fisher, Cohen, Holland, and Surtes (1986) and Ryan and Geiselman (1991) who used the so-called cognitive interview to assess memory, or McSpadden, Schooler, and Loftus (1988) who investigated the impact of context reinstatement procedures on test performance. Further examples include some work in connection with hypnosis (Sheehan & Grigg, 1985; Sheehan, Grigg, & McCann, 1984; Sheehan & Statham, 1989; Sheehan & Tilden, 1983, 1984, 1986; see Sheehan, 1988, for a review) and studies on child testimony (Howe, 1991; Rudy & Goodman, 1991; Saywitz, Goodman, Nicholas, & Moan, 1991; Warren, Hulse-Trotter, & Tubbs, 1991).

Finally, there are a few studies that have used various other measures of memory performance, sometimes in addition or in comparison to the procedures already discussed. Among these are reaction times (Cole & Loftus, 1979), implicit memory tests (Dodson & Reisberg, 1991), source monitoring tests (Lindsay & Johnson, 1989; Schumacher-Bittner, 1994; Zaragoza & Koshmider, 1989; see also part I), and multiple choice procedures (with or without a "don't know" option), the latter frequently being used in research on the impact of misinformation on person identification or -description (e.g., Franzen & Sporer, 1994; Gibling & Davies, 1988; Loftus & Greene, 1980).

So far, the eyewitness suggestibility landscape looks like that: early studies investigated - out of a mainly practical interest - the suggestibility of eyewitness reports as a function of variables of the questioning procedure itself, using a prototypical two-phase design with unspecific questioning cues intended to influence the subjects' estimates of a qualitatively different target variable. When aftereffects of such suggestibility manipulations had been discovered, research interests shifted to memory processes and to the impact of post-event misinformation per se. A first, intermediate step were *memory supplementation designs* (as we may call this kind of memory design) where the post-event misinformation merely supplemented the original event information; the question of interest was whether subjects would report this additional information when later tested for their memory of the witnessed event (mostly with a yes/no- or a recall procedure). The second step was the introduction of *memory contradiction designs* where the misleading details actually contradict details present in the original event. Corresponding to this logic, the typically used forced-choice recognition test (standard test procedure) sets the tendencies to report original or misleading details in opposition to each other. This type of design turned out to be the memory design par excellence, presumably because it allowed much more memory theorizing than is possible within a memory supplementation design: It does not make much sense to ask what happens to a memory representation of an event that is merely supplemented. However, it seems natural to ask what has happened to an original event detail when a discrepant post-event detail is reported instead.

As a consequence of asking memory questions, research done with memory contradiction designs has focused on typical memory variables like depth of encoding, time intervals, and retrieval cues, obviously at the expense of thinking about suggestibility factors. Is it really by chance that two influential studies which manipulated a suggestibility variable, namely, the credibility of the post-event information (Dodd & Bradshaw, 1980; Smith & Ellsworth, 1987)

did not employ a memory contradiction design but a supplementation design? Thus, it seems that research designs have set the stage for (or at least coincide with) a certain way of thinking about the phenomenon in question.

Particularly interesting from this point of view is the impact of McCloskey and Zaragoza's (1985) article mentioned above. When it became clear that suggestibility factors also contribute to subjects' performance on the standard test procedure, this was not taken to be important or interesting per se; the main point was that it questioned the suitability of the standard procedure with respect to the assessment of memory. Consequently, the primary concern of researchers has been how to get rid of these troublesome response biases: McCloskey and Zaragoza themselves proposed their modified test procedure (which was criticized for other reasons, see part I), and subsequently a full-sized debate regarding the proper (i.e., bias-free or bias-corrected) assessment of potential memory impairment emerged, in the course of which elaborate technical arguments were developed (see Belli, 1989; Loftus & Hoffman, 1989; Tversky & Tuchin, 1989; Zaragoza & McCloskey, 1989). This has certainly its merits[1]; on the other hand, however, it documents how far the two lines of eyewitness suggestibility research had drifted apart.

I will conclude this introduction into the eyewitness suggestibility research landscape with some orienting remarks on my own work. While my initial focus has also been on memory and its assessment, this led to the formulation of an integrative model of performance in eyewitness suggestibility experiments which, in a sense, can be seen as a reunification of the pure suggestibility line with the memory line in eyewitness suggestibility research. This is reflected in the two basic components of the model, a memory base and the transformation of this memory base into test performance. It is assumed that memory reports, though based on relevant knowledge in memory, also depend on the subjects' assumptions regarding possible answers to test questions which, in turn, are shaped by certain features of the test situation. The latter, of course, allows for the incorporation of suggestibility factors; the work presented in part I focuses on what I consider to be a major source of suggestibility in memory designs, namely, subjects' assumptions regarding the consistency of event and post-event information. A basic assumption of this two-component approach is that any attempt on a solution of the memory assessment problem must, in a systematic fashion, take both memory and suggestibility factors into account as well as specify how they interact with one another. This is a prerequisite for their possible separation for purposes of memory assessment. I criticize previous attempts (e.g., McCloskey & Zaragoza, 1985; Belli, 1989; Tversky & Tuchin, 1989) for having proposed only partial and ad hoc measures to correct for some biases which at some time became obvious with some existing procedures, instead of treating the problem as a whole.

Of course, it is left to the reader to judge whether my attempt has any advantages over those of my precursors. In any case, there is one benefit of a thorough conception: it allows for predicting new effects, like those investigated in part II of this work.

Paired-associate learning and interference

Paired associate learning is a technique developed for the experimental study of forgetting. The background for this approach (see Crowder, 1976, for an excellent review on which this introduction is largely based) lies in a commonly accepted view of the origin of forgetting which had emerged in the 1920s and the 1930s (summarized most clearly in a paper by McGeoch in 1932). This view contradicted the former belief that forgetting is due to the decay of memory traces, postulating instead that it resulted from the interference of later experiences, borrowing on the well-known phenomenon of retroactive inhibition ("rückwirkende Hemmung") which had been described by Müller and Pilzecker in 1900. Exactly how these later experiences interferred with existing memories, of course, was a matter of debate. Due to the behavioristic spirit of the time, the different proposals for interference mechanisms were cast in stimulus-response language: McGeoch himself put an emphasis on response competition (i.e., the idea that two responses associated to the same stimulus compete with each other at the moment the response is given), whereas Melton and Irwin (1940) held that some unlearning also must take place (unlearning means that the acquisition of the second response leads to an actual weakening of the first response's association to the stimulus).

The meaning of these concept becomes more readily apparent when we take a look at the design of a typical paired-associate learning experiment. Like a memory design in eyewitness suggestibility research, it consists of three phases: In the first, subjects are presented a list of so-called paired associates, that is, a stimulus term A (often a three-letter nonsense syllable) followed by a response term B (say, an adjective). This list is learned to a certain criterion (for instance, one perfect recitation). In the interpolated phase, then, the subjects learn a second list of paired associates. These consist of the same stimulus terms A as in the first list but paired with new response terms D. In the third phase, they are given the stimulus terms and are requested to recall the response terms from the first list. The typical finding is that experimental subjects (A-B, A-D arrangement) perform worse than control subjects who learn the same first list but a different second list, one that contains the same response terms but paired with new stimulus terms (i.e., an A-B, C-D arrangement). Of course, there are numerous variations of this basic design which are described elsewhere (e.g., Crowder, 1976).

Worth mentioning, however, is a variation of the recall test in the third phase which came to be known as the modified modified free recall (MMFR) procedure (Barnes & Underwood, 1959). This procedure resembles the test procedure I have used in some of my experiments. It

requests the subject to recall both the list 1 and the list 2 response associated with a given stimulus term. The rationale behind this procedure is that it was supposed to eliminate response competition, thus allowing for a pure assessment of unlearning. Indeed, Barnes and Underwood found considerable amounts of retroactive inhibition with the MMFR procedure; this was regarded as convincing evidence for unlearning.

Around 1960, it seemed that the two-factor (response competition and unlearning) theory of forgetting could provide a thorough explanation of retroactive inhibition and forgetting in general (this optimistic state is reflected in a review by Postman, 1961). Within a decade, however, this belief was seriously undermined after new phenomena and problems had been discovered. One of them was the so-called independent retrieval phenomenon (Greeno, James, & DaPolito, 1971; Martin, 1971). A cornerstone of the unlearning assumption had been the claim that unlearning is specific, that is, learning of a certain A-D combination in the second list leads to specific unlearning of the respective A-B counterpart from the first list. Now, Greeno et al. performed an itemwise analysis of the availability of response terms at test, distinguishing four possible outcomes which can be arranged in a 2*2 contingency table: For a given stimulus term A, either the B response or the D response, both of them, or none could be available to the subject. If unlearning was indeed specific, they argued, then the relative frequency of recalling both responses together should be lower than the product of the marginal frequencies with which any one of them is recalled (alone or together with the other; this logic exactly parallels my analyses in part I). What they found, however, was independence in the contingency table: retrieving a B response obviously was independent from retrieving the D response.

Another problem for the unlearning assumption was the almost complete absence of retroactive inhibition with recognition procedures (e.g., Postman & Stark, 1969): If the B responses really were unlearned during list 2 learning, then this should also lower performance on a recognition test. Finally, a major problem was how to account for proactive inhibition. If first-list responses were unlearned during second-list learning, how is it possible, then, that recall of the new responses was better in a control condition where no prior list was learned? The proposed solution to this dilemma was the concept of spontaneous recovery (Underwood, 1957): likening unlearning to extinction, it was held that there is spontaneous reappearance of extinguished responses after some time. Unfortunately, very little direct evidence for spontaneous recovery could be found (see Crowder, 1976). This latter difficulty was accounted for through the proposition of a response-set supression mechanism (Postman, Stark, & Fraser, 1968) which, however, was tantamount to a complete reformulation (or even abandonment) of the two-factor theory of forgetting. It was hypothesized that subjects, during list 2 learning, aquire a response set which has a certain inertia: Subjects tend, roughly speaking, to maintain their list 2 response set even when asked for the first list responses (this accounts for retroactive inhibition). However, this bias is only possible as long as the subjects are able to

distinguish the list 2 responses from the list 1 responses. This assumption pointed to an explanation for proactive inhibition: Since the ability of subjects to discriminate between lists decreases with time, the suppression of first list responses is no longer possible and performance for the second list suffers (for example, due to response competition). On the other hand, the new theory had its problems as well. For instance, there was no evidence for clustering of responses in free recall according to list membership (Martin, 1971), as might be expected if response sets are formed.

Thus, around 1970, interference accounts of forgetting were almost in a state of confusion (reflected in Postman & Underwood, 1973). It does not seem that this confusion has ever been resolved. Apparently, with the beginning cognitive revolution, paired associate learning became unattractive to many researchers, and they shifted to the new information processing paradigm which furnished them with new questions and research designs. As a consequence, the problems in paired-associate learning simply remained unsettled. However, while the theoretical framework was abandoned, the phenomena of retroactive and proactive inhibition in paired-associate learning designs have remained classics of research on learning and memory, and any modern theory of memory has to account for them.[2] From a more sarcastic point of view, on the other hand, one might note that paired-associate learning, initially a means for the study of forgetting, has now turned out to be an ends in itself, a development which had been foreseen and criticized as early as 1970 by Tulving and Madigan.

[1]However, it does not seem that a consensus on the memory assessment issue has been found; methods proposed as a solution to the problem (Belli, 1989; Tversky & Tuchin, 1989) have remained isolated episodes, and older procedures like the standard or the modified test procedure continue to be used.

[2]For instance, one of the most elaborate memory models, SAM (search of associative memory, e.g., Raaijmakers & Shiffrin, 1981), is able to simulate all of the major phenomena mentioned above (and others). Interestingly, Mensink and Raaijmakers's (1988) simulation of the Barnes and Underwood results with the MMFR procedure has been (successfully) done without the help of an unlearning assumption. Indeed, these authors consider their model to be a modern version of McGeoch's (1932) theory that emphasized response competition and changes in context between acquisition and retrieval.

Part I: Eyewitness testimony

I.1. Introduction to part I

When subjects are shown a simulated real-life event like a traffic accident, followed by a postevent account of the event containing some incorrect details, they tend to report these misleading details when tested for their memory of the original event. This has been labeled the *misinformation effect* (see for overviews, Loftus, 1991, Mélen, 1989; Sporer & aus dem Kahmen, 1994). Four major explanations for the effect have been advocated hitherto: (1) An alteration of the original event's memory representation (Loftus, 1979a; Loftus & Loftus, 1980; Metcalfe, 1990), (2) inaccessibility of original details (Bekerian & Bowers, 1983; Christiaansen & Ochalek, 1983), (3) response biases in favor of the misleading information (McCloskey & Zaragoza, 1985), and (4) misattributions of misleading details to the original event due to source confusions (Lindsay & Johnson, 1989).

According to the present state of knowledge, there seems to be empirical support for each of these accounts. This gives rise to the suspicion that the ongoing debate over the origin of the misinformation effect cannot solely be resolved on an empirical basis. Therefore, the first section of part I is devoted to an integrative framework which conceptually reconstructs the phenomenon and its explanations. This **I**ntegrative **M**odel for **P**erformance in eyewitness suggestibility experiments (IMP) also offers some methodological innovations for the study of eyewitness memory which are used in the following, experimental sections. Section I.3 (experiments 1 and 2) investigates potential memory impairment, using an unbiased and sensitive method suggested by IMP, while section I.4 (experiment 3) focuses on misinformation effects that arise from a deteriorated transformation of memories into test performance; thereby, a detailed account of subjects' processing of discrepancies is provided. Section I.5 combines both components in presenting a mathematical elaboration of IMP that simulates test performance on the basis of (a) frequency distributions of memory states and (b) unique features of the test procedure; this is exemplified by using the present experiments 2 and 3. An explorative section I.6 is directed at possible long-time effects of misleading information. Finally, a general discussion reviews the findings of the previous sections and relates them to research on eyewitness suggestibility in general.

I.2. The model and its theoretical and methodological implications

Most researchers agree that it is crucially important to distinguish between memory representations and memory reports in order to locate the effects of misleading information (e.g., Loftus, Schooler, & Wagenaar, 1985; Schooler & Tanaka, 1991). In the words of Zaragoza & Koshmider III (1989, p.254), "the key to explaining integration phenomena is understanding

the relation between the underlying memory representation(s) and subjects' performance on memory tests". The basic idea of the model presented here is to conceive of this relation as a *problem solving process*[1]: When answering memory test questions, the subjects are trying to find a solution to a] *memory task*. This solution is based on available information in memory but depends also on the subjects' *internal representation of the memory task*, which can deviate from its objective characteristics under certain circumstances, thus allowing for deteriorations of memory reports. This conception will now be worked out with respect to (a) a classification of aspects of memory representations which are relevant for solving the memory task and (b) the dependence of test performance on subjective assumptions of the consistency or non-consistency of information. The integrative and innovative power of the model is demonstrated by (c) reinterpreting existing accounts of the misinformation effect within this framework and (d) pointing out some methodological implications for assessing possible memory impairment.

To the reader's orientation, I want to note that I consider the conception of remembering as problem solving to be compatible with schema approaches in the Bartlett (1932) tradition, which also highlight the task-dependence and constructive nature of remembering (see Alba & Hasher, 1983, and Vandierendonck, 1986, for overviews). Just like Bartlett conceived of remembering as an adaptive function of the organism in order to meet the requirements of the environment, in IMP the subject tries to meet the requirements of the memory task, thereby using (or, temporally adapting) information stored in memory. However, Bartlett mainly had long-term processes in mind when thinking of adaptation (some of his subjects were asked to recount the "war of the ghosts" even after years), a perspective that will be taken up again in section I.6 and the general discussion.

Memory states

To consider which aspects of memory representations are relevant for solving the memory task, let us take a look at the typical design of eyewitness suggestibility experiments. In a first phase, the subjects are shown a simulated real-life event, for instance, a series of slides depicting a theft. Afterwards, they are presented misleading information concerning some critical details of the witnessed event, usually embedded in a narrative or a questionnaire. For other critical details, subjects receive no or neutral information (control condition). Third, subjects are administered a memory test which asks for certain details contained in the slides, for example: "What kind of tool was in the bag?". Now, availability of these original details is certainly relevant for solving the memory task, but also memory for the corresponding misleading details, because these may also provide a - false - solution to the memory task. Also, memory for the sources of the respective details is relevant, since the test asks specifically for the details from one source, that is, the initially witnessed event. Depending on the kind of memory test

used, knowledge about the sources of remembered details may influence the solution process. Now, we can build up a classification of *memory states* as follows (beginning with memory for an original detail):
I. Original detail available, source available.
II. Original detail available, source not available.
III. Original detail not available (availability of source irrelevant).
The same three memory states can be distinguished for the corresponding misleading detail. In the experimental condition, combination of these *simple* memory states for the original and misleading details yields nine possible *combined* memory states. In a sense, this classification of simple and combined memory states integrates and extends the considerations of McCloskey and Zaragoza (1985), who emphasized the role of memory for the misleading details, and Lindsay and Johnson (1989), who focused on source memory.
At a given point in time, an experimental subject occupies one and only one combined memory state. Which memory state a given subject occupies trivially depends on attention/perception processes (for instance, a subject may simply not notice all relevant details), and forgetting (unrelated to the impact of misleading postevent information). Furthermore, it might depend on memory impairment due to misinformation; this is an empirical question that can be adressed within the model.

The transformation of memory states into performance

How is the available information in a certain memory state used when a subject is solving a memory task?

Or, which part of it is used? IMP holds that the selection and use of available information depends on the *internal representation* of the memory task, that is, the subject's perception of the nature of the task and acceptable solutions. This internal task representation, in turn, is shaped by more or less conscious presuppositions or expectations. The role of subjective beliefs and expectations in remembering is also a standard topic of schema theories. Consider, as an example from research on autobiographical memory, the idea that *implicit theories* of people guide the recall of personal attributes (M. Ross, 1989). For instance, somebody entertaining the implicit theory that a study-skills course will help to improve his or her study skills might "remember" that his or her grades have been worse before the course, even though they remained stable over this period.

Within the context of eyewitness suggestibility experiments IMP draws attention particularly to subjects' beliefs regarding the *consistency* or *non-consistency* of presented information. This emphasis has to do with communicational aspects of psychological experiments which have come to the attention of researchers during the last years (e.g., Adler, 1984, Hertwig & Gigerenzer, 1994). With respect to eyewitness experiments, the point is that the common practice of deceiving subjects on the true purpose of the experiment and on the presence of misleading information - rather, for instance, they are told that the experiment wanted to find

---

---

out whether memory for the witnessed event would generally be better with a visual or a verbal mode of presentation (McCloskey & Zaragoza, 1985) - means a violation of the *cooperative principle* in human communication described by the philosopher H. P. Grice (1975), particularly a violation of maximes subsumed under the category of *quality*: In our daily communications, we tacitly assume that our counterpart does not say what he or she believes to be false or for which he or she lacks adequate evidence. Thus, in eyewitness suggestibility experiments, subjects will principally assume from the beginning (respectively, never think about the contrary) that the information presented in the second phase of the experiment is correct, that is, consistent with the information presented in the first phase. They will leave this assumption only if they have good reasons for doing so, for example, if they are told about the inconsistent details.

How will the subjects' internal representation of the memory task look like under these different assumptions? Usually, the memory test asks for specific details from a given source, that is, the original event. Now, under a consistency assumption it does not matter from which source a detail came, because the details are assumed to be identical anyway. Therefore, the memory task might be represented in the following way: "Search memory for the (one) relevant detail and respond with it". Conversely, when subjects know that some misleading details have been presented, the memory task is represented in a quite different way: "Search memory for relevant details. Then identify the source(s) of the detail(s) found. Respond with the detail from the slides".

A consistency representation of the memory task has important consequences on performance in the experimental condition whenever the subject has the opportunity to choose the misleading detail on the memory test, as, for instance, with the standard test procedure of Loftus, Miller, and Burns (1978). With a representation of the memory task as search for *one* piece of information, the search is stopped if any one detail (original or misleading, depending on the subject's memory state) is found; further searching must seem useless to the subject. Hence, IMP sees the acceptance of misleading details as test answers (McCloskey & Zaragoza, 1985) as the consequence of an inadequate internal representation of the memory task. Moreover, even if a subject can remember both the original and the misleading detail (having probably encountered a discrepancy), a consistency assumption might lead to a resolution of the conflict which denies the presence of two discrepant details (see section I.4). For instance, a subject might deliberate that the information provided by the experimenter is more reliable and therefore his or her own remembrances might be wrong (McCloskey & Zaragoza, 1985). In these cases too, experimental subjects' performance will be worse. In short, IMP holds that the well-known response biases connected with the standard test procedure are rooted in the subjects' consistency representation of the memory task.

Reconstructing existing accounts of the misinformation effect within the IMP framework

The impact of the subjects' consistency assumption is an inherent but unrecognized constituent not only in McCloskey and Zaragoza's (1985) account of the misinformation effect, as shown in the preceding paragraph, but also in the explanations of Christiaansen and Ochalek (1983) and Lindsay and Johnson (1989). The former postulated a systematic retrieval advantage of the misleading details because they were encoded more recently than the original event details. In order to demonstrate that the latter are still available in memory, Christiaansen and Ochalek employed a warning procedure to discredit the - according to their assumption - initially retrieved misinformation: They told their subjects prior to testing that the narrative contained misleading details, which should lead them to continue searching their memory and ultimately retrieve the original information. This argument implies that subjects' knowlegde about the presence of discrepant details enables them to access the original information, and, conversely, that the lack of such knowledge prevented (or at least helped to prevent) this access. Hence, Christiaansen and Ochalek implicitly draw on the subjects' consistency assumption in explaining the misinformation effect. Preferential access to misleading details is merely an additional assumption which is not necessary for an explanation of the misinformation effect, since the latter would also be obtained (though to a smaller extent) if the misleading information was accessed in some cases only. Thus, the reconstruction of Christiaansen and Ochalek's approach shows that it actually consists of two subarguments: subjects' consistency representation *and* preferential access to the misleading details.

Lindsay and Johnson (1989) attribute the misinformation effect to source confusions brought about by the standard test procedure which does not drive subjects' attention to the sources of the details. Such source confusions could be overcome using a new memory test which does drive subjects' attention to the sources. But, if we take a closer look, it becomes immediately clear that this memory test does far more than that. It requires subjects to indicate for each detail whether it has been presented (a) only in the slide, (b) only in the narrative, (c) in both of them, or (d) in none of them. Offering these alternatives, however, must lead subjects to the conclusion that there were indeed details which were not present in the slide[2], and has therefore the same effect as a warning: Subjects come to know that the information from the two sources is not consistent. Conversely, not paying attention to the sources with conventional memory tests seems reasonable on the basis of a consistency assumption, because they are assumed to contain the same information. Hence, it shows that the consistency assumption argument is also an implicit part of Lindsay and Johnson's approach.

To put the general argument more sharply: From the perspective of IMP, the various mechanisms invoked to explain the misinformation effect - responses biases elicited by the standard test procedure, blocked access to original event details, and source confusions enabled by the standard test procedure - are in fact consequences of the subjects' consistency representation of the memory task (as already noted, this holds only partially for Christiaansen and Ochalek's approach). This common core allows grouping these approaches together under the heading of]*transformational* hypotheses, a notion which emphasizes that the misinformation effect is not attributed to an impairment of the underlying memory base but to a deteriorated transformation of this memory base into performance, although each of the approaches emphasizes a different aspect of this transformation.

In turn, a second class of hypotheses can be distinguished which explain the misinformation effect in terms of *impairment*. This means usually a weakening/overwriting or reduced accessibility of the original memory trace (Loftus, 1991) but may also be extended to an impairment of source memory. These hypotheses can be implemented within the IMP framework as specific predictions for frequency distributions of memory states; this will be illustrated in the next paragraphs. - The two classes of hypotheses do not necessarily contradict each other but may correspond to different types of misinformation effects which may occur simultaneously in a given experiment; whether this actually takes place is an empirical question.

Methodological implications of IMP

There is one straightforward methodological implication of the dichotomization of existing accounts into transformational and impairment hypotheses: If you want to assess possible impairment, preclude transformational misinformation effects. Of course, this insight is not new; for instance, McCloskey and Zaragoza's (1985) modified test procedure precluded response biases simply by omitting the misleading details as possible responses. On the other hand, it is precisely because of this omission that the modified test procedure has been criticized for minor sensitivity to detect small degrees of memory impairment (Loftus, Schooler, & Wagenaar, 1985). According to IMP, this difficulty can be circumvented: The transformational misinformation effects that stem from the subjects' consistency assumption can be precluded by assuring a non-consistency representation of the memory task while still retaining the misleading details as possible responses (section I.4 will provide direct evidence for this claim). A suitable measure to ensure such a representation is the *enlightenment* of subjects on the true purpose of the experiment and on the presence of misleading details in the postevent account.[3] However, it must be assured the subjects are enlightened just prior to the test, that is, *after* the presentation

of misleading information. Otherwise, a possible integration might be prevented from the outset because the subjects would consciously look for discrepant details.

A second methodological benefit stems from IMP's theoretical classification of memory states. It offers the chance to compare empirical *frequency distributions* of memory states across subjects (the question how to identify these memory states will be discussed below) to predictions derived from different theoretical assumptions, for example, memory independence or impairment. At the same time, this permits to derive and assess the predictions from the various hypotheses at a more detailed level than usual in eyewitness suggestibility studies. For example, "normal" memory impairment can be clearly distinguished from source impairment.

The logic of the predictions will be illustrated first for the assumption of independent memory storage and retrievability; this is IMP's interpretation of the no-impairment assumptions implicitly or explicitly held by McCloskey and Zaragoza (1985) and Lindsay and Johnson (1989), for instance. We can say that across subjects the ability to remember an original detail (and its source) is independent from their ability to remember the misleading detail (and its source) if the relative frequency of any one combined memory state equals the product of the frequencies of the constituting simple memory states. Consider the hypothetical example given in figure 1. The combined memory state 1 refers to the ability to remember both details and attribute them to their respective sources. It is "made up" from the simple memory states I for the original and the misleading detail, where each detail per se can be remembered and properly located. Assume that 40% of the subjects occupy this simple memory state with respect to the original detail, and 50% with respect to the misleading detail. Then the expected frequency for the combined memory state in the experimental condition is 40% * 50% = 20%. The same logic holds for the other combined memory states.

There is a weak and a strong version of the independence claim. Weak independence simply means statistical independence in the 3 * 3 table; that is, the combined memory states are considered to be joint occurrences of two independent single events in the experimental condition (i.e., memory for the original detail and memory for the misleading detail). This way of defining independence is not new; it was independently developed some 25 years ago in the context of research on retroactive interference (Greeno, James, & DaPolito, 1971; Martin, 1971). A strong version of the independence assumption replaces the marginal frequencies in the experimental condition with the empirical (relative) frequencies of simple memory states in control conditions. The question is now whether memory for the original and the misleading details in the experimental condition is the same as if it had been synthesized from two control conditions, one for the details from the slide alone (as usual) and one for the details from the narrative alone (unusual). Note that an empirical investigation of strong independence requires to have two such control conditions (like in experiment 2), whereas an investigation of weak independence could be done without a control group. Also possible is a hybrid design (like that

--- 21 ---

used in experiment 1) which uses only the traditional control group for the original information.

Now, what follows if misinformation does impair memory for the original event?[4] Due to the impact of misleading information, it would be less likely for subjects in the experimental condition to remember the original detail. That means that there should be fewer combined memory states where subjects can remember both details; this refers to memory states 1, 2, 4, and 5 in figure 1. The expected degree of "loss" of these memory states depends on the assumed amount of impairment (x%), that is, there will be x % fewer of each of them. And what happens to the "lost" x%? Because in these cases subjects do not remember the original detail any more, they "turn into" the memory states 3 and 6, where subjects can remember the misleading but not the original detail. That is, the frequencies of these memory states expected under an impairment hypothesis are augmented by the "lost" x% of memory states 1, 2, 4, and 5 (see figure 1).

Misleading detail		Original detail: Detail + source	Original detail: Detail, no source	Original detail: No detail
		I 40%	II 20%	III 40%
Detail + source	I 50%	1 20% 15% 10%	2 10% 7,5% 5%	3 20% 27,5% 35%
Detail, no source	II 20%	4 8% 6% 4%	5 4% 3% 2%	6 8% 11% 14%
No detail	III 30%	7 12% 12% 12%	8 6% 6% 6%	9 12% 12% 12%

Figure 1. Frequency distributions of memory states (hypothetical example). Roman (arabic) numbers denote simple (combined) memory states. Within cells: top = independence, middle = 25% impairment, bottom = 50% impairment.

--- 22 ---

It is also conceivable that misinformation reduces subjects' ability to remember the *sources* of details. This is another possible understanding of the source confusion hypothesis (Lindsay & Johnson, 1989), which I want to refer to as the *source impairment* hypothesis, opposed to the already discussed claim that a certain test procedure simply does or does not drive subjects' attention to the sources of details. Now, source impairment can be thought of in many ways; for example, source memory for the original, the misleading, or both details may be impaired. Assuming the last, for convenience, this would mean that (some) memory states where one or both sources of the two details are available would turn into such where no sources are available, resulting in a specific deviation from independence, that is, a drift from the memory states 1, 2, and 4 towards memory state 5.

In order to compare such theoretically derived frequency distributions to empirical ones - which will be done in the following section I.3 - it is necessary that the latter be adequately measured. Therefore, two conditions must be met. First, subjects must be enlightened prior to the test in order to ensure an unbiased internal representation of the memory task, as already noted. Second, we need a memory test which can identify simple and combined memory states, that is, project different responses to a test question in a one-to-one fashion onto different memory states. The logic underlying such a test is simply asking for the constitutive elements of the memory states, that is, the details and their sources. Specifically, in the experimental condition the test asks for *both* details and their sources.

I.3. Investigating memory impairment

The two studies reported in this section employ IMP's methodological innovations to investigate the "classic" question whether memory for original details is impaired by misinformation. Previous attempts to answer this question were often criticized for using either biased or unsensitive test procedures (McCloskey & Zaragoza, 1985; Loftus, Schooler, & Wagenaar, 1985). Therefore, the aim of these studies was to assess memory with the unbiased and sensitive procedure suggested by IMP, thereby comparing obtained frequency distributions of memory states with predicted distributions assuming independence on the one hand or various degrees of memory impairment on the other hand. - Note that as a consequence of asking for both details and their sources in the memory test, more than two sources of information were needed in experiments 1 and 2 (a slide, a narrative, and a tape recording), because otherwise the identification of the sources of two details were trivial if at least one of them is known. However, as in traditional studies, subjects were presented only two discrepant details per critical item in the experimental condition. This offered the possibility to employ three different "discrepancy types": discrepancies between details contained in the slide and in the narrative, the slide and the tape recording, and the narrative and the tape recording. I considered this a good opportunity to see whether possible memory impairments would depend on the modes in

---

---

which the details are presented. - As already mentioned above, experiment 1 tested the predictions of strong independence only with respect to the original information, reflecting the conventional focus in eyewitness suggestibility research. This means that only the "traditional" control condition for the original details was run.

Experiment 1

Method

Subjects. 70 subjects (20 female, 50 male), mostly students from the University of Constance, were paid for participation.
Materials.In phase 1, one single slide was used; it was identical to the slide used by Lindsay and Johnson (1989) and depicted a cluttered office scene with four persons sitting or standing and a lot of objects placed on desks and shelves. The narrative was a text of approximately 400 words, essentially a german translation of the Lindsay and Johnson text except for changes in the critical items. The content of the tape recording was identical to the narrative except for the critical details. The text was recorded in the studio by a male speaker of a local radio station. There were two versions of both the narrative and the tape recording, corresponding to the requirements of the design.

Altogether, 12 critical items were used, partially overlapping with the original Lindsay and Johnson items. They are listed together with the alternatives used in the experiment and a third alternative (in parentheses) which solely appeared as additional response alternative in the memory test. Slide vs. narrative items: 1. The man's suit in the foreground: pin-stripe suit vs. plaid suit, (corduroy suit). 2. Clothing of man in the background: suit vs. cotton shirt, (pullover). 3. Object in man's hand: ball-pen vs. pair of scissors, (cigarette). 4. Color of man's hair in the background: dark vs. blond, (gray). Slide vs. tape items: 5. Facial hair of man in the foreground: moustache vs. beard, (no facial hair). 6. Clothing of woman in the background: police uniform vs. fire brigade uniform (office wear). 7. Glasses of man in the foreground: normal glasses vs. sunglasses (metal-rimmed glasses). 8. Drink on desk: cup of coffee vs. glass of orange juice (coke can). Narrative vs. tape items: 9. Heading of poster: Apollo 16 vs. Apollo 15, (Apollo 14). 10. Object on desk in the background: tape recorder vs. typewriter, (computer keyboard). 11. Accessoire on tie of man in the foreground: crime stoppers button vs. tie-pin, (monogram). 12. Lights on ceiling: 2 spotlights vs. 3 spotlights, (4 spotlights).

Design. Two critical items within each discrepancy type served as control items and two as experimental items (counterbalanced across subjects). In the control condition, the misleading details were simply omitted from the narrative or the tape recording. For the discrepancies between narrative and tape recording, four critical items were invented which did not appear in the slide. The assignment of critical details to sources was fixed rather than counterbalanced. For instance, "moustache" solely appeared in the slide, and "beard" solely in the tape recording.

Given the large number of critical items, I considered it unlikely that a possible misinformation effect would hinge upon specific combinations of original and misleading details.

Procedure. Subjects were tested in groups of 2 to 15. The study ostensibly wanted to find out how the succession of presentation of one and the same information in different modes (visual, written, and auditive) influenced memory for this information. - Subjects first viewed the slide for exactly 1 min. Then, after a 3 min unrelated filler task, they were handed out the narrative which they read at their own pace. After four minutes of reading, however, I reminded them to finish reading; the narratives were removed, and subjects took up again their filler task for another 3 min. Thereafter, subjects listened to the tape recording which lasted 2 min 50 sec for both versions. Immediately after the tape recording, subjects were enlightened on the true purpose of the experiment with approximately the following words: "Now I have a little surprise for you. What you read in the instruction sheet at the beginning of this session is not quite true. This experiment is about eyewitness testimony. I have inserted some discrepant details in the narrative and the tape recording; some of you may already have noticed that. I'm now interested in how you - as unvoluntary witnesses, so to speak - can remember these discrepant details. To this end, you will undergo a memory test ...", and so on. This enlightenment took about 2 min. Subjects then received the instruction for the memory test and were given three practice questions to become acquainted with the test. I demonstrated the correct answerings, and subjects had the opportunity to ask questions. Reading the instruction and working through the practice questions took about 12 min. Thereafter, subjects were given the memory test.

Memory test. The memory test consisted of 12 written questions, six for the control condition and six for the experimental condition. The order of the questions was randomly determined. Before each question for the control (experimental) condition the words "One detail" ("Two details") indicated that one (two) detail(s) had been presented; this should additionally support an adequate internal representation of the memory task, that is, as search for one or two details. The order of the three substantial response alternatives within each question was also randomly determined, whereas the order of the three source alternatives was always the same and corresponded to their temporal order in the experiment. For both the response alternatives and the sources, a fourth, "don't know" alternative was always on the lowest position. The test instructions reemphasized that misleading details had been inserted in the narrative and the tape recording and that there were three possible discrepancy types. Subjects were instructed to check the response alternative(s) and source(s) according to their memory. They were not encouraged to guess but to check "don't know" if they could not remember a certain detail or source. For experimental items, subjects were instructed to draw in lines when they had checked two details and sources each, in order to make their attributions discernible. Finally, for reasons to be discussed later, they were asked to indicate each question where they had noticed a discrepancy between the presented details previously in the experiment. - At the

end of the experiment, subjects were asked not to tell anything about the true purpose of the study to potential future subjects.

Analysis. Subjects' responses were analysed in the following way: Each properly checked response alternative or source was taken as evidence for the availability of the respective constituent of a memory state. For example, when a subject had checked "moustache" in a control condition question and had attributed it to the narrative, this indicated a simple memory state II (detail available, source not available), since "moustache" was the correct answer, whereas in fact it appeared in the slide. That is, any false (or "don't know") answerings were simply taken as evidence for the non-availability of a certain constituent of a memory state. There is a restriction to this logic in the experimental condition: When two alternatives and two sources were checked, but no connection was drawn in (in spite of the instruction to do so), the sources were dropped (i.e., counted as not available) because no clear attribution was discernible, even when the sources were the correct ones.

Results

Figure 2 gives the obtained frequencies of combined memory states and the predictions according to an independence assumption for all three discrepancy types. They are compared with each other using a goodness-of-fit test with a chi-square statistic; thereby, the independence predictions serve as the null hypothesis. Since experiment 1 investigated strong independence only with respect to the original details (i.e., no separate control condition for the misleading details was run), the frequencies of simple "misleading" memory states were estimated retrospectively from the respective marginal sums of the obtained frequency distributions of combined memory states. As a consequence of this estimation, two degrees of freedom are lost, so that the chi-square comparisons (based on N = 140 observations each) for each discrepancy type involve only six degrees of freedom. The predictions according to an independence hypothesis are derived following the logic outlined in section I.2, that is, by multiplying the empirical frequencies of simple memory states in the traditional control condition for the original details with the estimated frequencies of simple "misleading" memory states (numbers in parentheses).

The frequencies of combined memory states as obtained in the experimental condition show no significant deviation from these predictions for the "discrepancy types" slide vs. narrative ([chi]2 = 8.89) and slide vs. tape recording ([chi]2 = 12.19). A significant deviation from the independence predictions shows up for the discrepancy type narrative vs. tape recording
([chi]2 = 46.71, p < .001). Closer inspection of figure 2 reveals that this misfit is almost entirely due to the enhanced prevalence of memory states 1 and a corresponding decline in the frequency of memory states 2 and 4, compared with the independence predictions. Interestingly, these deviations are precisely in the opposite direction as would be expected from whatever version of a source impairment hypothesis.

To disentangle the impact of memory impairment in general from the impact of source impairment (or, enhancement) in particular, the data were also analysed without consideration of source memory. That is, memory states differing only in memory for the sources of details (those within the double lines in figure 2) were grouped together by simply adding their frequencies. Then, the same comparisons were performed as above (but with only two degrees of freedom remaining). Thus, any remaining deviations from independence solely reflect (potentially impaired) memory for the details themselves but not for their sources. This analysis yielded no significant deviations from the independence hypothesis for any discrepancy type (slide vs. narrative: [chi]2 = 0.53, slide vs. tape: [chi]2 = 0.09, narrative vs. tape: [chi]2 = 0.94; all df's = 2, all *N*'s = 140).[5]

For comparison, chi-square tests were also performed for predictions assuming various degrees of memory impairment. These predictions were derived as described in section I.2. Because of the already mentioned ambiguities of source impairment claims, this analysis was again done without source consideration. Since there is no amount of assumed impairment which were especially prominent in any way, I decided to calculate something like a "compatibility interval"[6], that is, a range of degrees of assumed memory impairment within which each value yields predictions that are (at the 5% alpha level with df = 2) statistically compatible with the actual data. This interval reaches from +17% to - 23% impairment for the discrepancy type slide vs. narrative, from +14% to -12% for slide vs. tape, and from +8% to -11% for narrative vs. tape. Thus, the data were compatible with negative amounts of impairment (i.e., memory enhancement) as well. The best fit (i.e., lowest chi-square) is obtained with -3% (slide vs. narrative), -1% (slide vs. tape), and -3% impairment (narrative vs. tape).

It is noteworthy that there was a considerable amount of reported discrepancy detection in the experimental condition of experiment 1. 47 out of 70 subjects had indicated that they had noticed at least one discrepancy between details presented to them (149 out of a total of 420 test answers). 126 discrepancies were reported when subjects could remember both details. 112 of these answers further corresponded to memory states 1, that is, when subjects could remember both details and properly attribute them to their sources.

a, slide vs. narrative

	ODS+	ODS-	OD-
Predicted:	I 47	II 5	III 48
MDS+ (57)	1 27	2 3	3 27
MDS- (26)	4 12	5 1	6 12
MD- (17)	7 8	8 1	9 8
Obtained:			
(57)	31	1	26
(26)	13	0	13
(17)	6	2	9

Figure 2. ***Predicted*** **and obtained relative frequencies of memory states in experiment 1 (roman/arabic numbers denote simple/combined memory states). Figures are percentages. ODS+ (-) / MDS+ (-) means the ability to remember the original / misleading detail and (but not) its source. OD- / MD- means that the original / misleading detail is not available. Figures in parentheses are the marginal frequencies (with respect to the misleading detail) of the obtained frequency distributions of combined memory states.**

Discussion

In general, the results are consistent with an independence assumption, with the exception of the discrepancy type narrative vs. tape. This deviation, however, is restricted to source memory, as can be concluded from the fact that no similar deviation is found when not considering source memory. On the other hand, the results are incompatible with any serious amount of memory impairment (or, enhancement). Therefore, the main conclusion to be drawn from experiment 1 is that misleading information did not impair subjects' ability to remember original

information. This result is in line with the arguments and evidence provided by McCloskey and Zaragoza (1985) and others. In addition, however, the present results cannot be criticized because of minor sensibility of the test procedure to detect possible memory impairment (Loftus, Schooler, & Wagenaar, 1985) or memory blends (Metcalfe, 1990). Unlike the modified test procedure used by McCloskey and Zaragoza, the present test procedure suggested by IMP includes the misleading details as response alternatives. Therefore, if memory for the original details were impaired, subjects had the opportunity to act according to this impairment and choose the misleading detail on the test.

A restriction to these conclusions might be that the present results were obtained along with a considerable amount of discrepancy detection. It has been argued (Loftus, 1979b; Tousignant, Hall, & Loftus, 1986) that discrepancy detection prevents possible impairment of original memory to occur. If this argument holds, the present results would be rather uninformative with respect to memory impairment. However, there is evidence that the large number of indicated discrepancy detections in experiment 1 is not an isolated result; there are other studies where similar numbers were found with different materials and different procedures (Geiselman, Fisher, Cohen, Holland, & Surtes, 1986; Ryan & Geiselman, 1991; Pohl, Schumacher, & Friedrich, 1993). Thus, discrepancy detection is quite possibly a pervasive phenomenon in eyewitness suggestibility experiments, including studies where a misinformation effect was obtained. Of course, this would cast the preventive power of discrepancy detection into doubt (though we cannot finally decide about that because most studies do not report any information pointing to this issue). Indeed, section I.4 provides direct evidence that discrepancy detection does *not* automatically lead to resistance against misinformation.

Moreover, a coincidence of discrepancy detection and good memory for the original information does not necessarily mean that the latter is a consequence of the former. After all, good memory for the original (and the misleading) detail is to some degree a *precondition* for detecting a discrepancy. Had I not encoded the original detail, I would not notice a discrepancy to the misleading detail. Hence, discrepancy detection might be a mere by-product of good memory for the discrepant details and might not enhance memory for them at all. Still another possibility is that discrepancy detection drives subjects' attention to the sources of the details ("Oh, I think here in the narrative there was something quite different from the slide!"), which results in better encoding and, consequently, better retention of the sources; this will be termed the *source attention hypothesis*.

Post-hoc evidence for this hypothesis comes from the discrepancy type narrative vs. tape. The remarkable deviation from the independence predictions in the upper left fields of the respective 3 * 3 table in figure 2 (less memory states 2 and more memory states 1 than predicted under an independence assumption) coincides nicely with the large number of discrepancy detections in this discrepancy type: out of the 78 (56%) memory state 1 answers, there were 48 answers with discrepancy detection. Thus, it is imaginable that in some of these cases subjects

who otherwise would have forgotten the source of the original information (ending up in memory state 2) could keep it in memory (yielding memory state 1) because the detection of a discrepancy drove their attention to it and gave it an additional rehearsal. - Independent evidence for this interpretation of the effects of discrepancy detection on source memory comes from a recent study by Pohl et al. (1993).

To summarize, the results of experiment 1 are consistent with an independence assumption and with the hypothesis that discrepancy detection enhances source memory. Moreover, the former holds, in essence, independently of the mode of information presentation (visual, written, auditive). Nevertheless, it should be kept in mind that experiment 1 was able to test the strong independence assumption only with respect to the original details. It may be worthwile to see if strong independence also holds with respect to the misleading details. Therefore, experiment 2 attempted a replication of experiment 1 with a design that provides separate control conditions for both the original and the misleading details. Moreover, additional changes were introduced to investigate two potentially critical features of experiment 1: First, by indicating for test questions in the experimental condition that two details had been presented, subjects in experiment 1 were possibly provided a strong cue to work hard in searching their memory for these two details. It could be argued from the perspective of the accessibility hypothesis (Bekerian & Bowers, 1983; Christiaansen & Ochalek, 1983) that the otherwise poorer accessibility of the original details had been overcome by such a measure. Therefore, experiment 2 abandoned such cues and left it to the subjects to decide whether they had been presented one or two details. Second, in experiment 1 it was not possible to decide about guessing tendencies which could have contaminated the estimated frequencies of memory states (a good estimation was needed for the simulations to be reported in section I.5). Therefore, experiment 2 was also designed to control for guessing tendencies.

Experiment 2

Method

Subjects. 73 subjects (43 female, 30 male) were paid for participation, mainly students from the University of Salzburg engaged in various studies; further, four *Gymnasium* pupils and one university lecturer participated. Subjects were aged from 17 to 62 years (*Md* = 21). They were tested in groups of size 1 to 8.

Design. Three separate control conditions for the critical details from the slide (C1), the narrative (C2), and the tape recording (C3) were run. In each of these conditions, no reference to the critical details was made in the other two sources of information, that is, the respective details were omitted from the slide, the narrative, or the tape recording. Additionally, a guessing control condition (C4) was run where subjects received no information at all concerning a test question. There were three experimental conditions corresponding to the three possible

---

---

discrepancy types (i.e., E1: slide vs. narrative, E2: slide vs. tape recording, and E3: narrative vs. tape recording). All of these conditions were run within subjects. The assignment of critical items to conditions was counterbalanced across subjects according to a latin square; this yielded altogether 8 counterbalancing groups of subjects. The experimental condition C2 (slide vs. tape recording) was run twice because the design did not work with only seven conditions.

Materials. Eight fairly heterogenous critical items were used, covering a wide range of saliency or centrality (the critical details used in the slide, the narrative, and the tape recording, in this order, are added): A. Potted plant: kind of cactus, sedge, Yucca; B. Cigarette brand: Camel, Marlboro, Gauloises; C. Pamphlet heading: Konstanz (Constance), Mainau (Isle of Mainau), Bodensee (Lake of Constance); D. Sign heading: Zimmervermittlung, Zimmerreservierung, Zimmerservice (three roughly synonymous terms meaning accomodation service). E. Form of glasses: oval, round, rectangular; F. Object of customer: cotton bag, camera bag, plastic bag; G. Head-covering of customer: cap with shade, hat, woollen cap. H. Poster heading: Blasorchester (brass orchestra), Streichorchester (string orchestra), Kammerorchester (chamber orchestra). As in experiment 1, the assignment of critical details to sources of information (i.e., slide, narrative, and tape recording) was fixed. - The altered design required two different versions of a slide which did or did not contain each four of the eight critical details. Both versions are shown in figure 3 (together with a third version which was used in experiment 3). The slide depicts the interior of the Tourist Information office, Constance. The two versions were constructed on location by adding or dropping from the scene, respectively, four details each when the pictures were taken. The narrative in phase 2 was an approximately 420-word-long description of the slide. There were 8 versions of this narrative, each containing 3 critical details according to the requirements of the design. The tape recording was identical to the narrative except for the critical details; it was recorded by a male radio speaker in the studio. There were two versions of the tape recording, each one containing four critical details.

Procedure and memory test. The procedure was the same as in experiment 1 except for some temporal changes: slide (1 min) - unrelated filler task (12 min) - narrative (self-paced, about 3 min) - second filler task (12 min) - tape recording (about 3 min 20 sec in both versions) - third filler task (5 min) - enlightenment (about 2 min) - memory test. The memory test consisted of 8 questions concerning the critical details. Unlike experiment 1, subjects now had two separate but otherwise identical response options for each question. Within each option, they had to choose between the critical item versions listed above and "don't know" and attribute their choice to one of the sources (also including "don't know"). The second response option was headed by the phrase: "If you think that you have got two pieces of information:". Accordingly, the test instructions emphasized that subjects use this second option if they remembered two details. It was maintained that, if they had received two details, these could be discrepant or consistent. The reason for this false assertion (with respect to the experimental items asked for in the memory test) was the following. If subjects remembered one information and knew

that they had received a second, which, however, they could not remember exactly, they had a better guessing chance if they knew that the second information always differed from the first. This, in turn, would mean an undue advantage for the experimental conditions over the control conditions and might contaminate the results. Further, it was not mentioned that there was one question on the test where they had received no information (i.e., condition C4). This was done to yield a fair measure of guessing tendencies. - Reading through the instructions took about 1,5 min, after which subjects immediately took up the memory test. Because of the changed features, the memory test was easier to handle, so that the practice questions could be dropped.

Analysis. In addition to the analysis of memory states (as in experiment 1), *response patterns* for conditions were compared. These were meant to give a detailed overall description of subjects' rememberings in a certain condition. Since the assignment of critical item versions to sources was fixed, subjects' responses in a certain condition can be aggregated across test questions. Thus, combining three detail alternatives (associated with the slide, the narrative, and the tape recording, respectively) and three source alternatives (augmented by the "don't know" option, respectively) yields sixteen possible *response types*, which are chosen with a certain frequency; these frequencies can be represented in a 4 * 4 table. This is henceforth called a response pattern. Note that subjects' answers from *both* of the two separate response options for each question entered the same 4 * 4 table. As a consequence, the total relative frequency of responses in a response pattern could easily exceed 100% and maximally reach 200%.

Prior to entering the response pattern or memory state analyses, subjects' responses were processed in the following way. Each response was fully counted except when a subject had checked the same detail twice (mostly attributing it to two different sources); in this case, each response was counted half because in fact the detail was remembered only once and fully counting each response would lead to an overestimation of the frequency of this response type, and, consequently, of memory for this detail. Responses were excluded from analysis if two response alternatives and two sources were checked within the *same* response option, or if two details were attributed to the same source. However, such cases were rare. Additionally, there was a frequency correction for both the response pattern and memory state analysis. To account for detail characteristics and unequal group sizes (there were 8 - 10 subjects in each counterbalancing group), relative frequencies were first calculated within one group (remember that each counterbalancing group had a different critical detail in a given condition), and in a second step the mean of these item-and-group-specific relative frequencies was calculated to yield the overall relative frequency of a certain response type; the same principle holds for the calculation of the frequencies of memory states.

--- 32 ---

Figure 3. Slide versions used in experiment 2 (top and middle) and experiment 3 (bottom).

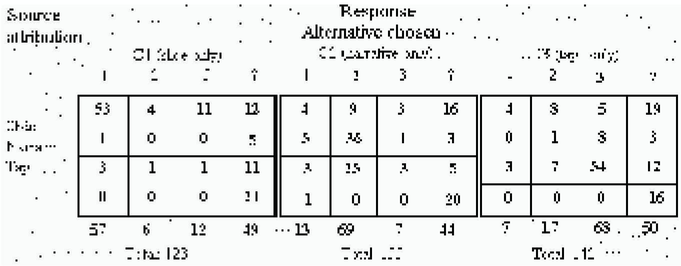
--- 33 ---

Results

The response patterns for all conditions are given in figure 4. All response patterns except for E2 are based on 73 or 72 (due to exclusions) observations. The response pattern for E2 is based on 145 observations (one exclusion) because this condition was run twice. Direct comparisons between control and experimental conditions regarding the original information (i.e., with respect to retroactive interference) can be made for C1 vs. E1, C1 vs. E2, and C2 vs. E3. 57% of the subjects remembered the original information (regardless of source attribution) in C1, vs. 56% in E1. Comparing C1 and E2 yields 57% vs. 52%, and the values for C2 vs. E3 are 69% vs. 69%. The respective comparisons for the misleading information (i.e., with respect to proactive interference) yield 69% in C2 vs. 71% in E1, and 68% in C3 vs. 68% in E2 vs. 61% in E3. McNemar tests for the differences between these proportions indicate that none of them is significant (all chi-squares < 1). The same holds if we compare the frequencies of those response types where the original or the misleading detail is properly attributed to its source.

Figure 5 gives the empirical frequency distributions of memory states for the three experimental conditions[7] together with the predictions of an independence hypothesis. Comparisons between the independence predictions and the obtained values yielded][chi]^2 (8, N = 73) = 11.97, p > .1 for E1, [chi]^2 (8, N = 145) = 37.64, p < .001 for E2, and [chi]^2 (8, N = 72) = 30.42, p < .001 for E3. Like in experiment 1, the data were also analysed without consideration of source memory. These comparisons yielded no significant deviations from independence: [chi]^2 (3, N = 73) = 3.04 for E1, [chi]^2 (3, N = 145) = 3.13 for E2, and [chi]^2 (3, N = 72) = 1.50 for E3.[8] For comparison with memory impairment predictions, again compatibility intervals were calculated (again, without source consideration). These revealed that the results were compatible (alpha = 5%, df = 3) with amounts of memory impairment ranging from 40% to -15% in E1, from 37% to -8% in E2, and from 26% to -22% in E3. The best fits were obtained for 13% (E1), 15% (E2), and -2% (E3) impairment. - Unlike experiment 1, this time it was also possible to run the same analyses with respect to an impairment of memory for the]*misleading* details, that is, with respect to proactive inhibition. The respective predictions are derived in the same way as for

"normal" memory impairment, with the one difference that the x% "impaired" memory states add to those memory states where only the original details are available. These analyses yielded compatibility intervals from 35% to -13% in E1, from 26% to -11% in E2, and from 34% to -15% in E3. The best fits resulted for 9% (E1), 6% (E2), and 7% (E3) proactive impairment.



Source attribution	C1 (slide only)				C2 (narrative only)				C3 (tape only)			
	1	2	3	?	1	2	3	?	1	2	3	?
Slide	53	4	11	12	4	9	3	16	4	8	5	19
Narrative	1	0	0	5	5	36	1	3	0	1	8	3
Tape	3	1	1	11	3	15	3	5	3	7	54	12
?	11	0	0	21	1	0	0	20	0	0	0	16
	57	6	12	49	13	69	7	44	7	17	68	50
	Total 123				Total [illegible]				Total [illegible]			

Figure 4. Response patterns in experiment 2. Numbers are percentages. The numbers 1, 2, and 3 in the top rows refer to the details presented as response alternatives within a test question ("?" means "don't know"). When presented as critical details in the respective condition (bold numbers in larger fonts), the "1" details always appeared in the slide (respectively, 2 = narrative, 3 = tape).

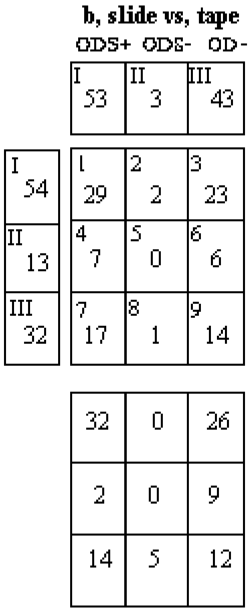


Figure 5. ***Predicted*** **and obtained relative frequencies of memory states in experiment 2. Numbers are percentages. ODS+ (-) / MDS+ (-) means the ability to remember the original / misleading detail and (but not) its source. OD- / MD- means that the original / misleading detail is not available.**

The number of detected discrepancies was even larger than in experiment 1. Only three subjects did not report any discrepancy detections. The relative frequency of detected discrepancies in the experimental conditions were 51% (E1), 52% (E2), and 59% (E3), respectively. With respect to memory states, 94 noticed discrepancies (out of 290 answers in the experimental conditions) were indicated when subjects could remember both details. 84 of these answers further corresponded to memory states 1, that is, when subjects could remember both details and properly attribute them to their sources.

Discussion

The frequency distributions of memory states in the experimental conditions show essentially the same pattern of results as in experiment 1. When taking source memory into account, there are some significant deviations from the independence hypothesis that vanish if we abandon source memory and concentrate on the details themselves. These deviations are mainly drifts from predicted memory states 2, 4, and 5 (i.e., perfect detail memory but suboptimal source memory) towards memory state 1 (perfect detail and source memory). They are interpreted

like in experiment 1, namely, as tentative evidence for the source attention hypothesis, that is, the assumption that detecting a discrepancy between details leads to better encoding and retention of the sources of these details.

Unlike experiment 1, the results indicate that this time some (retroactive) memory impairment might have occurred, though the overall amount would be rather small (some 10% across discrepancy types). In comparison with experiment 1, this could be interpreted as tentative evidence for an accessibility hypothesis which states that the now missing information about the number of presented details provides less optimal retrieval conditions for the original information. However, this cannot account for the similar, though somewhat smaller, amount of *proactive* impairment for the misleading details. On the contrary, the assumption of minor accessibility of the original information is built upon the assumption that there is *preferential* access to the misleading details. From the perspective of IMP, the symmetrical nature of the retroactive and proactive inhibition in experiment 2 can be accounted for without assuming memory impairment, namely, as another type of a transformational misinformation effect: Remember that the internal representation of the memory task is assumed to be influenced by subjective expectations. In the test instructions, subjects were told that they had received one or two pieces of information for each question. Consequently, they probably assumed that they had received 1.5 details per question on average. Now, unless memory is perfect for all test questions, it is likely that in a number of cases subjects will mistakenly think that they have received two details on a control question and, in order to arrive at the envisaged average quantum of 1.5 details, received only one piece of information on an experimental question. Evidence for this sort of regression toward the mean comes from the response patterns: In the control conditions C1-3, subjects gave roughly 1.3 answers, though, since subjects had received only one piece of information, these additional answers necessarily had to be either fabrications (i.e., choices of false response alternatives) or "don't know"-responses. Conversely, in the experimental conditions subjects gave roughly 1.7 answers for each test question, imaginably because they did not search their memory as thoroughly as if they had known about the true number of presented details. Furthermore, they might even have rejected some uncertain memories.

A further change to experiment 1 was the addition of a guessing control condition. The response pattern for C4 reveals that there were virtually no guessing tendencies in this condition even though subjects had been told that they had received at least one piece of information per test question (however, subjects followed the instruction insofar as they provided roughly 1.1 responses in this condition on average, mostly "don't know" responses). Furthermore, some of the few substantial responses might be mistaken recognitions of slide details, that is, reflect no guessing at all. Thus, the obtained frequencies of memory states in conditions C1-3 and E1-3 are most probably not contaminated by guessing tendencies and can be regarded as good estimators of the actual frequencies of memory states. This holds for experiment 2 (neglecting the small "impairments" in the experimental conditions) and, in turn, also for experiment 1.

The central message to be taken from the results of experiment 2 is that misleading information essentially had no effect on the ability of subjects to remember originally presented information. Insofar, experiment 2 replicated the results of experiment 1 under somewhat harder conditions. That is, the results of experiment 1 hold up if we derive the predictions of the independence hypothesis for performance in the experimental condition from separate control conditions, and they essentially hold up if we do not tell subjects about the exact number of details they have been presented. As a whole, the results of these two studies using the methodology inspired by IMP imply that subjects could simultaneously remember two discrepant pieces of information as well as they could remember each of them alone, with the exception of better source memory in the experimental conditions, due to discrepancy detection. - It should be noted that in a recent study done by Belli, Lindsay, Gales, and McCarthy (1994), who had independently employed a procedure which comes very close to the one used in experiments 1 and 2 (i.e. asking for both details together with previous enlightenment; however, Belli et al. used a recall procedure), better control than experimental memory for original details had been found. The authors interprete this finding as evidence for memory impairment. There is no room here to discuss their interpretation at length. Assuming its validity (though, one may adopt an alternative explanation discussed by the authors, which is similar to the regression argument advocated above), the validity of the present interpretation of the results of experiments 1 and 2 might be restricted to recognition procedures. This would be in line with older findings from the domain of paired-associate learning, where retroactive inhibition was constantly found with recall but not with recognition procedures (e.g., Crowder, 1976).

I.4. Investigating transformational misinformation effects

While the preceding section was occupied with the memory base in eyewitness suggestibility experiments, this and the next section deal with the transformation of the memory base into performance. In section I.2, I have pointed out that this transformation is mediated by the subjects' consistency or non-consistency assumptions. Particularly, I have argued that two causes of the misinformation effect which are commonly associated with the standard test procedure - response biases (McCloskey & Zaragoza, 1985) and source confusions (Lindsay & Johnson, 1989) - are in fact consequences of the subjects' consistency assumption. If this is true, then it should be possible to overcome the influence of misinformation even with the standard test procedure, simply by ensuring a non-consistency representation of the memory task, that is, by enlightening the subjects. Thus, experiment 3 compared two groups of subjects: One group (the standard group) was run as usual, whereas the other group (the enlightenment group) was enlightened prior to the memory test; the usual manipulation of presenting or not presenting

misleading information was run within these groups. Now, if (non-) consistency assumptions are the crucial variable, then there should be no misinformation effect in the enlightenment group, paralleling the non-effects of misinformation obtained in experiments 1 and 2, whereas a misinformation effect should emerge in the standard group.

A second point of interest in this experiment was the impact of discrepancy detection, which is a rather neglected issue in research on eyewitness memory, as already noted. A notable exception is some work of Loftus and her coworkers (Loftus, 1979b; Tousignant et al., 1986). Essentially, these authors assume that discrepancy detection reduces suggestibility. On the other hand, from the perspective of IMP it is conceivable that the subjects' consistency assumption will also dominate the processing of discrepancies and reduce or preclude possible resistance effects. In any case, knowledge about that may deepen our understanding of the "inner mechanisms" of suggestibility. Since I already expected from experiments 1 and 2 that subjects would detect a lot of discrepancies, I considered this a good opportunity to learn about subjects' processing of the discrepant details, and planned a detailed qualitative analysis of subjects' "coping strategies" for discrepancies, as we might say.

Experiment 3

Method

Subjects. 53 subjects took part in the experiment, seven of whom were excluded from analysis for various reasons.[9] The remaining 46 subjects (24 in the standard group and 22 in the enlightenment group) were undergraduates and graduates of the University of Salzburg, Austria; 20 were psychology students (10 in the standard group and 10 in the enlightenment group). Additionally, one teacher took part in the experiment. Subjects were aged between 18 and 50 years (]*Md* = 22) and were paid for participation.

Materials and design. The materials used were the same as in experiment 2, except for two changes due to the design of experiment 3: (1) The slide used in phase 1 now contained all eight critical details (see figure 3), and (2) no narrative was presented to the subjects. That is, all subjects saw eight critical details in the slide and received misleading information for four of these details from the tape recording; in the control condition, they received no information for the respective details. The assignment of critical items to the control and experimental condition was counterbalanced across subjects.

Procedure. Subjects were tested in groups of size 1 to 7. They were assigned to the standard or enlightenment group according to a predetermined schedule. The procedure was the

same for both subject groups except that the enlightenment group was enlightened just before the test and the standard group only after the memory test. For reasons which will become clear in section I.5 the procedure also paralleled experiment 2 except for changes due to the design of experiment 3. That is, because subjects were not given a narrative, the filler task between the slide and the tape recording had to be longer (about 28 min). Equally, the time interval between the end of the tape recording and subjects beginning to work on the memory test was the same as in experiment 2 (about 10 min). This was ensured by adjusting the length of the second filler task in both the standard group and the enlightenment group.

Memory test. The memory test consisted of 16 written two-alternative forced-choice questions. The first eight questions were filler questions. The last eight questions referred to the critical items, with the response alternatives always being the respective original vs. misleading details. The order of the two response alternatives (indicated by preceding letters A and B) was counterbalanced across subjects. The instructions maintained that across all memory tests given to subjects in the experiment, the response alternatives A and B were equally often correct. For individual memory tests, they were told, this was not necessarily true (though in fact it was). Therefore they should not try to count how often they had already checked A or B but base their responses on what they had seen in the slides. This latter phrase was made especially salient through bold type. This rationale seemed to me a suitable way of instructing subjects to remember the details from the slide without suggesting all to openly that there might be different details in the tape recording.[10] When subjects had read through the instructions, they were given the memory test. Answering the test questions took about 4 or 5 min on average. Subjects in the standard group were now enlightened, and both groups were given a last sheet where the topics of the test questions and the response alternatives had been re-listed; subjects were then requested to indicate any discrepancies they had noticed and to describe as accurately as possible the nature of these discrepancies and what came to their minds when experiencing them. Additionally, they were requested to describe how these deliberations had influenced their responses on the memory test.]

Results

Table 1 shows a clear misinformation effect in the standard group, that is, a significant difference between control and experimental performance, *t* (23) = 3,64, p < .001, one-tailed. There is no misinformation effect in the enlightenment group. Furthermore, the equal levels of

performance in the control conditions of each group indicate that there were no systematic differences between the groups in their ability to remember critical details per se. - The frequency of discrepancy detection was comparable to experiment 2. Again, only three subjects did not report any discrepancy detections. In the experimental condition of the standard group, there were 48 discrepancy detections (50% of the answers), compared to 40 (45%) in the enlightenment group. Analysing the discrepancy detections with respect to the truth value of the corresponding test answer (see table 2) reveals a fundamental difference between the standard and the enlightenment group. Whereas in the latter virtually no false answer was given when a discrepancy was detected, there was a substantial amount of false answers in the standard group. There is a similar difference for answers where no discrepancy was detected; I will return to this issue below.

Table 1: Percent Correct Recognition in Experiment 3

	Standard Group (N = 24)		Enlightenment Group (N = 22)	
Experimental	50	(31 - 69)	78	(58 - 91)
Control	78	(58 - 90)	78	(58 - 91)

Note. The figures in parentheses are the 95% confidence intervals (calculated after Hays, 1973, p. 379).

Table 2: Number of Correct and False Responses in the Standard and Enlightenment Group separated after Discrepancy Detection (only Experimental Condition)

	Standard Group		Enlightenment Group	
Discrepancy	Correct	False	Correct	False
Detected	28	20	38	2
Not detected	20	28	31	17

Note. The total number of responses is 96 (88) in the standard (enlightenment) group.

The subjects' retrospective accounts of their discrepancy detections were highly different in completeness of detail but altogether sufficiently precise to allow further processing. Following common content-analytic procedures (e.g., Holsti, 1969; Mayring, 1988), I have analysed them with respect to three major categories which were particularly prominent in the subjects' accounts or interesting from a theoretical point of view: Statements concerning (A) memory for the original detail, (B) the interpretation or resolution of the discrepancy, and (C) the response strategy on the memory test. This was done for every single discrepancy detection. Several typical patterns of coincidences emerged; I borrow the term *process histories* from Loftus and Hoffman (1989) to refer to these different ways a given test answer could have come about. I restrict myself to the main process histories; a more detailed account, including methodological issues, is given elsewhere (Blank, 1994). Table 3 provides an exhaustive assignment of discrepancy detections to process histories, together with the numbers of correct and false answers associated with these process histories. Let us first turn to the process histories in the standard group.

1. Doubting own memory: The most salient feature of this process history was that subjects casted their own memory of the original detail into doubt (or were unsure about it, or thought they hadn't paid attention to that detail) when experiencing a discrepancy to the misleading detail; however, subjects typically reported that their memory for the original detail was not good from the beginning. Consequently, they accepted the misleading detail at the memory test. - This process history is mainly identical to the deliberation mechanism described by McCloskey and Zaragoza (1985).

2. Interpretation: This process history - or class of process histories - is best characterized through the fact that subjects could resolve the experienced discrepancy in a way that accounted for the presence of two discrepant details. Further, good memory for the original details prevailed and was the basis for subjects' (correct) responses on the memory test. There were three types of such interpretations/resolutions:

(a) Interpretation as mistake. Here, subjects assumed that the speaker of the tape recording had inaccurately observed or named details from the slides. Interestingly, subjects entertained this interpretation only for one detail at a time (at least within this sample). Moreover, this interpretation was compatible with process history 1. That is, the same subjects could generally doubt their own memory but interprete one single discrepancy as a mistake.

(b) Interpretation as method. Some subjects assumed that the presentation of discrepant details simply was a method of finding out whether subjects would base their responses more on the visual or the auditive presentation, an interpretation which indeed came very close to the true purpose of the experiment, without, however, implicating that the experimenter had misled the subjects on that.

(c) Interpretation as deception. Here, subjects took the detected discrepancies as evidence that misinformation had been presented and the pretended purpose of the experiment was not the true one. Two subjects reported that this deliberation had led them to scrutinize the rest of the tape recording for further discrepancies, as Loftus (1979b) had assumed. A third subject is an interesting exception insofar as she first seemed to engage in the first process history (doubting own memory) but in the last moment revised her interpretation and her response

strategy. I want to give a somewhat longer quote from this subject because it shows - like in a microcosm - the operation of processes which normally do not take place within one and the same person:

```
(Statement referring to discrepancy detection at detail C; subject is quoting
her own thoughts word by word:) "Strange, I only saw pamphlets with the heading
'Bodensee', maybe I didn't notice the others." (Detail A:) "... this wasn't a
Yucca after all, wasn't it? Maybe I fooled myself." (Detail G:) "... obviously
these filler tasks rather confused me, but I believe this wasn't a woollen cap.
Well, my memory is quite a bugger; fortunately, there is also this auditive
information, otherwise I probably would have forgotten everything." (General
```

```
statement:) In the instruction it struck me that it was emphasized to trust
one's own (subject's emphasis) memory for the slide. Only then I had the
idea that there might have been misleading information, and I looked for the
question with the Yucca. There my assumption was confirmed. Also, I was sure by
then that 'Bodensee' had indeed been written on the pamphlets.
```

3. Sure memory: This process history subsumes cases where subjects merely indicated that they were sure about the original detail and/or had clearly recognized the discrepancy, without, however, providing an interpretation. Consequently, these subjects chose the original details on the memory test.

Table 3 shows that the process histories in the enlightenment group are similar to those in the standard group. However, there is an important difference in the first process history. Due to the enlightenment prior to the memory test, subjects now re-trusted their memory for the original details which they had cast into doubt just before. Consequently, they based their responses on this re-gained confidence and virtually gave no false answers. - There was a comparable number of subjects who reported an interpretation of the discrepancies as mistake, method, or deception. Also, some subjects merely indicated that they were sure about their memory for the original details while at the same time giving no information referring to an interpretation of the discrepancies. Unlike the standard group, there were two subjects who merely reported a resistance to further misleading information, that is, that they had been consciously looking for further discrepancies to the slide.

Some more general remarks of the subjects also point to their response strategies for answers where they did not detect any discrepancies: Some subjects in the standard group reported that they generally trusted the tape recording and based their responses on it (corresponding to misinformation acceptance as conceived by McCloskey and Zaragoza, 1985). In contrast, a few subjects in the enlightenment group reported a response conflict when they could only remember the details from the tape: They were not sure whether they should accept or reject these details because they did not know whether they were true or not (indeed, the tape information for the filler questions was correct). That is, enlightenment did not simply lead to a reversed bias (i.e., rejection of post-event information).

Process History	Correct	False	Subjects[b]
		Standard Group	
1. Doubting own memory	1	12	10
2. Interpretation as			
a, mistake	6	--	6
b, method	4	--	3
c, deception	7	--	3
3. Sure memory	10	--	6
4. Others	--	6	2
Unclassifiable[a]	--	2	1
Sum	28	20	
		Enlightenment Group	
1. Doubting and re-trusting	8	1	6
2. Interpretation as			
a, mistake	4	--	3
b, method	6	1	3
c, deception	6	--	3
3. Sure memory	5	--	3
4. Resistance	5	--	2
Unclassifiable[a]	4	--	2
Sum	38	2	

Table 3: Exhaustive Classification of Discrepancy Detections in Experiment 3 into Process Histories

Note. The figures in the left and middle column are the numbers of correct and false responses associated with the respective process history. Figures in the right column are the numbers of subjects contributing these responses. See text for details.

[a]Unclassifiability resulted from subjects' providing too little information to extract a process history.

[b]The numbers of subjects within groups do not add up to the total group size because the same subject could be contributing responses to different process histories for different items.

Discussion

The overall result of experiment 3 supports IMP's claim that transformational misinformation effects are heavily dependent on the consistency or non-consistency assumptions of subjects

but are not produced by the standard test procedure itself: There is a sizeable misinformation effect in the standard group but a null effect in the enlightenment group; this is exactly what one should expect, because the procedure of experiment 3 paralleled that of experiment 2, where no misinformation effect was found either.

Another striking result of experiment 2 is that there was a large misinformation effect in the standard group *in spite of* the high rate of discrepancy detection. This means that discrepancy detection *per se* does obviously not lead to resistance against misinformation, as Loftus (1979b) and Tousignant et al. (1986) assumed. Rather, as plausible from the perspective of IMP, its effect is mediated by the consistency or non-consistency assumptions subjects hold. This is revealed by two comparisons: First, within the standard group, susceptibility to misinformation was found with process history 1, where the interpretation or resolution of the discrepancies obviously implied the assumed presence of only one, consistent information, but not with process history 2, which accounted for the presence of two, discrepant details. Second, subjects in the enlightenment group virtually gave no false responses associated with process history 1, while the opposite is true for the standard group. This difference can only be attributed to the enlightenment manipulation which destroyed subjects' consistency assumption that had led them into doubting their memory before.

On the other hand, we may ask if the non-consistency assumption associated with the "resistant" process history 2 in the standard group was not the *consequence* of the previous discrepancy detection. This would mean that, at least in some cases, discrepancy detection does in fact lead to resistance (and, as a by-product, to a non-consistency assumption). The picture becomes clearer if we take two further mediating factors into account, original memory strength and the number of detected discrepancies. It turns out that in the standard group resistance to misinformation following discrepancy detection was associated with good prior memory, with an average of roughly two discrepancy detections per subject (process histories 2 and 3), whereas susceptibility to misinformation was associated with poor prior memory and roughly one discrepancy detection (process history 1).

These interrelations may be expressed and resolved in a simple process model of resistance/susceptibility to misinformation: The subjects start with a consistency assumption and try to hold this assumption as long as possible, that is, as long as the number of discrepancy detections is low or memory for the respective original details is poor. The encountered discrepancies are interpreted according to this consistency assumption, and subjects deliberately opt for the misleading details. With more discrepancies detected or better original memory, it gets more difficult to retain the consistency assumption. At an intermediate stage, subjects may interprete one clear discrepancy as a mistake, which can be seen as a local exception to the general consistency assumption and allows resisting that particular piece of misinformation. Eventually, however, a threshold is crossed and the consistency assumption is left, maybe after the second clear discrepancy. This means stable resistance against misinformation and, furthermore,

leads to a reinterpretation of earlier discrepancies (remember the quote from the subject). Enlightenment, of course, bypasses this "natural" threshold (the precise location of which presumably depends on personal characteristics like ambiguity tolerance or on previous experiences with psychological experiments, for instance).

This model, even though it may be an overgeneralization from the present results, provides a nice account for the susceptibility of subjects to misinformation in previous eyewitness suggestibility studies, even if there was a considerable amount of subjects who detected discrepancies. Since most studies employ only one or two misleading details, it is unlikely that these subjects will cross the resistance threshold. Therefore, it almost logically results that there will be no overall resistance to misinformation.

I.5. Mathematical simulation of transformational misinformation effects

While the last section dealt with more qualitative aspects of the transformation of memory into performance, this section presents a mathematical elaboration of IMP which is a quantitative approach to the central problem stated in section I.2: understanding the relation between memory representations and subjects' performance on memory tests. The logic of the approach is quite simple: Based on IMP's classification of memory states, test performance in the control or experimental condition can be subdivided into performance contributions of single (simple or combined) memory states. Then, total test performance is the weigthed sum of performance within single memory states; the weights are the relative frequencies of the memory states. Mathematically, this is expressed by the following formulae:

$$\text{Total control performance} = \sum_{i=I}^{III} f_{ci} * P_{ci} \quad \text{and}$$

$$\text{Total experimental performance} = \sum_{i=1}^{9} f_{ei} * P_{ei},$$

where f_{ci} (f_{ei}) is the relative frequency of a simple (combined) memory state in the control (experimental) condition, p_{ci} (p_{ei}) is the expected single performance within this memory state, and I - III (1 - 9) are the running numbers of the simple (combined) memory states. The single performances within memory states have to be determined individually for each memory test. In a sense, the model follows the same logic as McCloskey and Zaragoza's (1985) numerical examples that served to illustrate the origin of poorer experimental performance on the standard test procedure. However, the memory base is much more complete, particularly with respect to source memory, which is of major importance for performance within memory states (see below).

The application of the model is best illustrated using an empirical example; at the same time, this is a first test of the model. Experiments 2 and 3 of the part I were explicitly designed to allow for such a test. Remember that experiment 2 was intended to give an estimate of the frequency distribution of memory states at the time of test and that the temporal course of experiment 3 was parallel to experiment 2. Therefore, the frequency distributions of memory states in experiment 2 (see figure 5, slide vs. tape) can be taken as a basis for predicting performance in experiment 3. However, this will only be done for the enlightenment group, since the simulation of performance in the standard group would involve additional parameters (for instance, for suggestibility) which would make the simulation very complex. Figure 6 shows the expected single performances for the standard test procedure used in experiment 3, together with previous enlightenment.

Experimental performance:

	Detail + source	Detail, no source	No detail
Detail + source	1 100%	2 100%	3 pR1
Detail, no source	4 100%	5 50%	6 pR2
No detail	7 100%	8 1-pR2	9 50%

Control performance:

Detail + source	Detail, no source	No detail
I 100%	II 1-pR2	III 50%

Figure 6. Expected performances within memory states (standard test with enlightenment).

In the experimental condition, let us first consider the performances within the memory states 1, 4, and 7, where the original detail can be remembered and properly located. Even if the misleading detail is also remembered (memory states 1 and 4), there should be no reason for

subjects to choose the latter on the test because of the previous enlightenment. Therefore, the performance within these memory states is assumed to be 100% correct. The same holds for memory state 2, because in a traditional eyewitness experiment with two sources of information the subject can infer the (not remembered) source of the original detail from his or her memory for the misleading detail and its source. [11] Also, the expected performances for the memory states 5 and 9 pose no problems. In the former, subjects can remember both details but not their sources. Therefore, they have to guess when asked which of the details was presented in the slides and, consequently, have a 50% chance of being correct. With memory state 9, the subject remembers nothing and has to guess anyway.

In memory state 3, all the subject remembers is the post-event detail and its source. Whether to accept or reject this detail therefore depends on the subject's knowlegde about the truth value of post-event information in general. Extreme cases are the lack of enlightenment, which would lead to misinformation acceptance (cf. McCloskey and Zaragoza, 1985), and telling subjects that there was only false post-event information with respect to the test questions (as Lindsay, 1990, did), which should lead to complete misinformation rejection. However, the mere knowlegde that there are]*some* misleading details in the post-event account (which mirrors the situation in experiment 3) constitutes an intermediate case. Here, the final decision whether to reject or accept the remembered detail should depend on subjects' beliefs concerning the relative amount of misleading details. These beliefs, however, are hard to specify exactly. Therefore, I have employed a parameter pR1 to denote the probability that the remembered detail will be rejected on the test. Simultaneously, this is the probability of being correct on the test (because it is the misleading detail that the subject rejects, and with the standard test procedure this means choosing the correct alternative).

In the memory states 6 and 8, the subjects remember a certain detail they had encountered during the experiment but not its source. Here, they should feel much less inclined to reject such a detail than in memory state 3 where they at least know that it came from the narrative and they should be cautious. Therefore, the rejection probability pR2 in the memory states 6 and 8 should be markedly lower than pR1. - While pR2 is assumed to be equal for memory states 6 and 8, the expected performance is not. This is simply because in memory state 6 the subjects remember the misleading detail and are therefore correct when they reject this detail on the test; the opposite holds for memory state 8 where subjects remember the original detail.

Deriving the expected single performances in the control condition is fairly trivial for the simple memory states I and III. In the former, the subjects remember the original detail and its source; hence, they will choose it on the test. In the latter, they know nothing and have to guess. With memory state II, the subjects are in the same position as with memory state 8 in the experimental condition: They remember a certain detail, which happens to be the original detail, but not its source. Therefore, the same arguments apply as above, and performance is assumed to be identical to memory state 8.

Performance in experiment 3 can now be simulated assuming different values for the decision parameters pR1 and pR2. For the experimental condition, not the obtained empirical frequencies of combined memory states are used as the basis of the simulations, since these are somewhat contaminated through a "regression effect" (remember the discussion of experiment 2). Instead, the predicted frequencies according to an independence assumption are taken (see figure 5, top 3 * 3 table). The frequencies of simple memory states in the control condition are estimated from C1. The empirical results (78% correct in both the control and the experimental condition) are best fitted when pR1 is 50% or 60% and pR2 is zero or close to zero.[12] These rejection parameters correspond well to the above considerations and also to subjects' response strategies (without discrepancy detection) reported in section I.4. Furthermore, pR2 is considerably lower than pR1, as was argued above. The good fit of the simulations is an indirect support for the model and, moreover, for the validity of the procedure employed in experiment 2 for the detection of memory states, since the logic of the simulations rests upon the assumption that the frequencies of the memory states have been adaequately estimated. On the other hand, it should be mentioned that values of pR1 between 40% and 70%, for instance, yield simulations within five percent points of the empirical results (with pR2 ranging from zero to (pR1)/2). Therefore, this is certainly no hard test for the validity of the IMP simulations.

In principle, this kind of simulation can also be extended to multiple choice or recall tests or to biased procedures (i.e., without enlightenment). Surely, this requires an individual adaptation of the within-memory-state performances. Moreover, instead of post-hoc simulations on the basis of empirical frequency distributions of memory states, it is also possible to simulate performance a priori from hypothetical frequencies of memory states. This may support theoretical analyses of the properties of different test procedures. One may ask, for instance, how performance on these tests will vary with different underlying memory bases. Interestingly, such simulations demonstrate the possibility of another kind of misinformation effect which is neither due to memory impairment nor to consistency assumptions but a result of decreasing source memory (not source impairment!) and certain properties of the test procedure (this will be elaborated in Part II). With the standard test procedure, for example, such effects are caused

by source confusions in memory state 5 and subjects' tendency to accept single remembered details with unknown sources (memory states 6 and 8) as answers.

I.6. Long-time effects of misleading information?

This more explorative section is concerned with memory performance after a long time interval. A theoretical interest in such performance comes from the abovementioned IMP simulations: Since it is very plausible that source memory fades with time, there should be ample opportunity for such effects to occur. Although, according to the simulations, they should be smaller in magnitude than the usual effects due to a consistency assumption, I was interested to see if they could be empirically confirmed. Moreover, a practical interest in long time intervals is simply that they do occur very often in eyewitness testimony. Typically, there is a delay of months before a case is put to trial. Even though the police try to question eyewitnesses immediately after the incident, there may be additional issues not captured by the initial testimony. - To investigate this, I had the simple idea to test my subjects from experiment 2 and 3 once again after some months, hoping that source memory would have faded substantially in the meantime. That is, in a follow-up study, I mailed the subjects the same test questions as in the experiment.

Follow-up study for experiments 2 and 3

Method

Subjects. The subjects were a subset from those participating in experiments 2 and 3. From the 73 subjects of experiment 2, 54 responded (5 to 8 subjects in each counterbalancing group). 28 of the 46 analysed subjects from experiment 3 responded, 16 from the standard group (8 in both counterbalancing groups) and 12 from the enlightenment group (6 in both counterbalancing groups). The subjects were not paid for participation. Instead, ten times 50 ÖS (4 $) were disposed of by lot among the respondents as an incentive.

Design. See experiments 2 and 3, with one exception: Because the standard group in experiment 3 had also been enlightened (*after* the memory test), the constituting difference between standard and enlightenment group did not exist any more.

Procedure. Letters were sent to the subjects four and a half months after experiment 2 and 3, respectively, and they were asked to answer the test questions. Because participation in the follow-up study was voluntary, I kept the instructions short. Subjects in the follow-up study for experiment 2 were reminded (1) to decide for each test question whether they had received one or two details and (2) to check for any remembered detail one of the alternatives on the left side and one of the sources on the right side (or "don't know"). Correspondingly, in the follow-up study for experiment 3 the subjects were reminded to (1) base their responses solely

on their own memory for the slide, (2) that they had received discrepant information in the tape recording for some details, and (3) that they definitely had to choose one of the test alternatives.

Memory test. In the follow-up study for experiment 2, the memory test consisted of the same eight questions as the original test in the experiment. However, the order of the questions was changed (i.e., once more randomly determined). Just as the original order, the new order was not correlated with the temporal order in the tape recording. Also, the order of the response alternatives was changed (with the "don't know" alternative staying in the fourth position). In the memory test for the experiment 3 follow-up, four of the filler questions at the beginning of the questionnaire were dropped in order to save subjects' time. As for the experiment 2 follow-up, the order of the questions and of the response alternatives was changed.

Results

Due to the explorative character of this follow-up study and to save space, the results are reported in not so much detail than for the experiments themselves. However, they were analysed in exactly the same way. Figure 7 shows the response patterns for all conditions. They reveal some marked differences to the response patterns of experiment 2 (see figure 4). First of all, there is much more "noise", that is, many more false alternatives were chosen. This is most obvious in condition C4, where subjects had received no information at all. In experiment 2, virtually no false alternatives had been chosen in this condition. Second, there was a pronounced tendency to attribute the chosen alternatives (including these fabrications) to the slide. Third, while the number of correct answers remained stable for the details from the slide, it declined for the details from the narrative and from the tape recording. However, this decline was less pronounced in the experimental conditions than in the control conditions. Finally, memory for the sources of the details was much worse than in experiment 2; this was especially salient for the details from the narrative and the tape recording. - The analysis of the frequency distributions of memory states requires no separate tables; basically the same pattern of results emerges as in experiment 2. The 3 * 3 tables (i.e., with consideration of source memory) show significant deviations from the predictions made according to an independence assumption, for all experimental conditions except E3: $[chi]^2$ (8, *N* = 54) = 20.18, $p < .025$ for E1; $[chi]^2$ (8, *N* = 54) = 26.87, $p < .001$ for E2; and $[chi]^2$ (8, *N* = 53) = 14.93, $p > .05$ for E3). However, the 2 * 2 tables (i.e., without source consideration) show no significant deviations from independence: $[chi]^2$ (3, *N* = 54) = 5.33 for E1, $[chi]^2$ (3, *N* = 54) = 7.33 for E2, $[chi]^2$ (3, *N* = 53) = 1.11.

	E1 (slide vs. narrative)				E2 (slide vs. tape)				E3 (narrative vs. tape)			
	1	2	3	?	1	2	3	?	1	2	3	?
Slide	39	23	10	8	45	15	20	2	16	31	11	3
Narrative	4	12	3	9	5	5	18	3	5	11	17	4
Tape	1	11	5	3	5	5	9	3	9	10	16	2
?	7	7	4	22	2	6	7	26	2	2	4	35
	51	53	22	41	56	31	54	24	32	54	47	43
	Total: 167				Total: 176				Total: 176			

[Figure 7]

The results of the experiment 3 follow-up were collapsed over the former standard and enlightenment group because this difference was now obsolete. Performance was 63% correct in the control condition and 61% correct in the experimental condition. An IMP simulation of these results on the basis of C1 and C3 in figure 7, assuming independence and rejection probabilities pR1 = 50% and pR2 = 0%, yielded predictions of 80% control and 70% experimental performance. For a second simulation attempt (see discussion), a guessing correction procedure was employed; this involved subtracting the whole response pattern C4 cell by cell from pattern C1 and from pattern C3. The frequencies of simple memory states were then taken

from these corrected patterns. With the corrected frequencies, the same simulation as above yields 64% control and 61% experimental performance.

Discussion

Let us first turn to the experiment 2 follow-up. A striking feature is the increased prevalence of noise. Whatever its origin may be (I will return to this issue below), it is present in all conditions. Thus, it is highly probable that the numbers of correct responses in the conditions C1-3 and E1-3 are an overestimation of true memory for the respective details. An indirect support for this claim comes from the fact that an IMP simulation of the performance in the experiment 3 follow-up on the basis of the uncorrected response frequencies yields predictions that are too high. Only when the noise is removed (by subtracting the "pure measure" of noise, C4, from the other response patterns) a good fit can be achieved. The corrected frequencies, on the other hand, reveal that a good amount of forgetting since experiment 2 did take place. This holds for the details themselves as well as for their sources. Apart from these overall trends, there is no compelling evidence for any differential impairment or enhancement of memory in the experimental conditions compared to the control conditions. With respect to the deviations from the predictions of the independence assumption in the 3 * 3 tables, the same arguments can be applied as in experiment 2.

Now I want to return to the noise phenomenon. This could be interpreted as mere guessing. However, equally plausible (and theoretically as well as practically important) seems to me another possibility: These wrong answers could represent a real misinformation effect, which, however, would have occurred in *every* condition (not only in the experimental conditions). Remember that the subjects had encountered these details twice *after* their "actual" presentation in experiment 2, namely, in the memory test and in the discrepancy detection questionnaire (where the question topics and response alternatives had been listed once again). Hence, it is quite plausible that some subjects remembered these additional, "unactual" presentations of details. Further, as was the case in general, they might have forgotten the sources of these details and thought that they had encountered them somewhere in the "actual" experiment. This belief would be reinforced in the follow-up memory test, since the subjects had no opportunity to indicate an unactual origin of a piece of information (simply because I did not think of this problem when planning the follow-up study).

However, this argument does not explain why the noise details were attributed especially to the slide. This remains an open question. It could be that, due to the richness of detail, it is easier to visualize a certain detail in the slide than in the two other sources, or that the subjects simply reasoned that there were more details in the slide, so that it is more probable that a certain detail came from the slide. On the other hand, there is the possibility that this pronounced attribution to the slide reflects an *integration* of the unactual, additional details into the memory representation of the slide, as Loftus (1979a) has proposed. Such an integration would be

even more puzzling because it would have occurred in spite of the previous enlightenment and without the details being embedded in a description of the slide. Whatever the reason is, the pronounced attribution of "remembered" details to the slide may be a phenomenon in its own right and, moreover, of enormous practical importance.

The results of the experiment 3 follow-up do not show a significant misinformation effect. However, if we take the results at face value (i.e., statistical considerations let aside), they show a minimal misinformation effect which is in line with the (corrected) IMP simulation. The small magnitude is due to a floor effect: Memory was generally poor, and so there were only few memory states 5 (both details but not their sources available) and 6 (misleading detail but not source available) which are thought to be mainly responsible for a misinformation effect of this kind. Apart from that, it is worth noting that the same IMP simulation (i.e., using the same rejection probabilities) produced a good fit for the original experiments 2 and 3 as well as for their follow-up studies. Although these are no hard tests, as already discussed, IMP did at least not fail them.

I.7. General discussion of Part I

Let us briefly recapture the major findings of the previous sections. According to the integrative model presented in section I.2, the theoretical accounts for the misinformation effect can be reduced to two main classes: impairment hypotheses and transformational hypotheses. It was argued and shown (in section I.4) that the transformation of memory into performance can be heavily deteriorated by consistency assumptions (but there are other transformational misinformation effects like the "regression effect" found in experiment 2 or the source memory effect suggested by the simulations in section I.5). In contrast, no evidence for memory or source impairment was found when examining the underlying memory base in section I.3 (the claim that this base was reliably captured with the methodology suggested by IMP was underpinned by the good fit of the simulations done in sections I.5 and I.6). A major side-issue was the impact of discrepancy detection on memory and performance. It was found in section I.3 that discrepancy detection enhances source memory for the discrepant details, and section I.4 showed that it does not produce resistance against misinformation unless a threshold is crossed where the discrepancies can no longer be reconciled with the subjects' consistency assumption. Finally, section I.6 added a further dimension to the misinformation issue in discovering that long-time processes and effects of post-event information may be quite different from those operating at an immediate level. Particularly, the misrememberings found after months were unrelated to the misinformation manipulation done in the actual experiment but seemed to stem from the response alternatives presented in the initial memory test. Moreover, the long-time tendency to attribute (mis-) remembered details to the slide may reflect their integration into a generalized representation of the experiment. - Now, I discuss these issues

with respect to their relation to and their implications for research on eyewitness suggestibility. This refers to (1) the impact of (non-) consistency assumptions on memory performance, (2) memory impairment, (3) the transformation of memory representations into memory reports in general, and (4) integration phenomena and their relation to source memory and discrepancy detection.

Non-consistency assumptions and non-effects of misinformation

The results of the present studies are in line with an increasing body of research demonstrating non-effects of misinformation on prior memory; a substantial part of this research was done with the modified test procedure of McCloskey and Zaragoza (see Belli, Windschitl, McCarthy, & Winfrey, 1992, for a review of these studies). However, particularly interesting from IMP's point of view are some other studies employing procedures that served to enhance a non-consistency assumption of the memory task (Christiaansen & Ochalek, 1983; Dodd & Bradshaw, 1980; Kroll, Ogawa, & Nieters; 1988; Kroll & Timourian, 1986; Lindsay & Johnson, 1989; Smith & Ellsworth, 1987); this has already been shown for two of these studies in section I.2. With respect to the others: Dodd and Bradshaw (1980) told their subjects that a description of a traffic accident was provided by a neutral bystander vs. the lawyer of the driver who obviously caused the accident (experiment 2), or the driver himself (experiment 3). While a misinformation effect was obtained in the neutral bystander condition, this was not the case when subjects were informed that the account came from a biased source. Smith and Ellsworth (1987) presented postevent information which either came from a student who had seen the event in question several times (i.e., a competent observer) or only once for a short time (incompetent observer). A misinformation effect showed up in the first but not in the second condition. In both studies, the subjects had reasons to assume that the maximes of quality (Grice, 1975) were violated: it was likely that the driver or his lawyer would manipulate the account of the accident in their own favour, that is, say things that they believe or know to be false, and it was likely that the incompetent observer's report lacks adequate evidence. Consequently, they could not assume the consistency of this information with the original event. In two further studies (Kroll & Timourian, 1986; Kroll, Ogawa, & Nieters, 1988), a misinformation effect established in an initial memory test was diminished in a second test prior to which the subjects were told about the presence of misleading details (however, both studies also involved a context reinstatement manipulation which contributed to the overall results). Moreover, this second test contained the misleading details as possible responses. Hence, the procedure of these two studies was similar to my own studies, and so were the results. Altogether, these four studies corroborate the findings in this work and lend further support to IMP's claim that memory reports are strongly sensitive to assumptions of consistency or non-consistency of information.

Relative contribution of memory impairment and transformational misinformation effects

The studies cited above as well as the findings in this work suggest that memory is not impaired by misinformation. However, some evidence for memory impairment does come from studies where the difference between event and postevent information is evident to the subjects or they are tested with the modified test procedure. I do not want to discuss these studies in detail; at present, though, it seems that there are two types of such impairment effects: First, there are small (i.e., roughly a ten percent difference between experimental and control performance) and temporally restricted impairment effects that can be found in studies where subjects are tested more or less immediately (i.e., up to half an hour) after the presentation of misinformation (e.g., Belli et al., 1992, 1994; Chandler, 1989, 1991, 1993), paralleling older findings from research on retroactive inhibition (e.g., Postman, Stark, & Fraser, 1968). It seems fair to say that such effects are negligible for practical purposes or at least not very dramatic. Second, there are somewhat larger and apparently stable impairment effects which, however, were only found with very young children (Ceci, Ross, & Toglia, 1987). Apart from these limited impairment effects, however, the bulk of research on eyewitness suggestibility seems to reflect transformational misinformation effects, that is, adaptations of memory to the requirements of the task or the situation. The importance of such effects has been increasingly recognized in the field (e.g., Loftus, 1991); they should be studied in their own right instead of asking the memory impairment question time and again. For instance, it would be interesting to see under which conditions outside the laboratory response biases occur (a study by

Wagenaar & Loftus, 1990, suggests that eyewitness identification procedures do frequently induce such) and, perhaps even more interesting, how they can possibly be avoided (for example, is there an equivalent of enlightenment or warning outside the laboratory?).

A decision on this topic would require a theoretical clarification of the processes that govern the transformation of memory representations into memory reports in various situations in general; response biases are but one way of adapting memory to situational demands. IMP's account of this issue might be a step in this direction, though it certainly has to be complemented and refined (since it was developed to account for performance in eyewitness suggestibility *experiments* in the first place). For example, Moston (1990) reviewed four sources of child suggestibility in eyewitness interviews, some of which seem familiar from the above discussions: the source credibility of the misleading information, demand characteristics, repeated questioning and the linguistic form of questions. The last point has been the topic of a considerable body of research (for instance, Davis & Schiffman, 1985, Kallio & Cutler, 1987, Lipscomb, McAllister, & Bregman, 1985, Loftus & Zanni, 1975). A more radical conception of remembering *as* discourse can be found in Edwards and Middleton, (1986a, 1986b). In line with this approach, Bekerian and Dennett (1990) investigated the impact of communicative

conventions on memory reports (i.e., spoken and written recall). Other candidates for transformational accounts are the schema approaches already mentioned in section I.2. Schuurmans and Vandierendonck (1985) combine the schema view with conversational issues: In their work on recall as communication, they show that the recall of objects in an office (or, ommission from recall) depends on the anticipated knowledge of a hypothetical recipient (in one condition, "martians who have never been on earth"). Finally, research on the influence of social context on memory may be relevant to the transformation of memory representations into memory reports (for instance, Higgins & Stangor, 1988; Holtgraves, Srull, & Socall, 1989).

Integration without impairment

Now, while memory impairment, especially trace impairment, seems to be rare in eyewitness suggestibility experiments, what about integration? This question seems odd at first glance, since the integration claim was hitherto always associated with the notion of trace impairment (Loftus, 1979a, Loftus & Loftus, 1980). However, this is no necessary conjunction. Integration denotes the incorporation of postevent details into a pre-existing memory representation of a whole event (Loftus, 1979a), whereas trace impairment (i.e., overwriting or blending) is held to take place at the level of single details (Loftus & Loftus, 1980). Hence, while trace impairment logically includes integration, the reverse is not the case. Integration might simply involve supplementing the event representation with additional details, for instance, if the respective event detail has been forgotten or never encoded. Moreover, this might even occur when there is still an original detail in the representation that contradicts it. The remembering person then would hold two memories he or she associates with the event in question but could not decide upon the true one (why not hold discrepancies in memory?). In any case, it seems plausible that such integration would be associated with a loss of source memory for the added details as well as for the original details possibly retained; it would not be much of an integration if the remembering person still knew that a certain detail did not come or did clearly come from the original event.

However, there are two ways to conceive of this association: The loss of source memory may either be the consequence of or the precondition for integration. The former would mean source impairment, the latter not. If the latter was true, then integration could be the end-product of a kind of selection process: a non-event detail the source of which was forgotten or never encoded could be retrieved because it is a plausible answer to a question. Then, as a consequence of this retrieval, this detail (a) received additional rehearsal and (b) was associated with the event. In turn, the likelihood of a further retrieval in association with the event increases. Thus, over some time, the non-event detail could be gradually integrated into the event representation; maybe this is what Loftus & Hoffman (1989) had in mind when speaking of the creation of new memories, and what Bartlett (1932) meant with the adaptive function of remembering.

Note, however, that this would be a rather long-time process, instead of the more

spontaneous process which the trace impairment view of integration depicts. Such an "evolutionary" integration may have caused the large amount of misinformation attributed to the slide in the follow-up of experiment 2 (section I.6).

The art of integrating discrepant information without detecting discrepancies

In any case, whether there is spontaneous or evolutionary integration, a major obstacle to it would be the detection of discrepancies between possible candidates for event details, as was hypothesized by Loftus (1979b) and Tousignant et al. (1986). The results of section I.3 suggest a possible mechanism for that: Integration (not a misinformation effect! see section I.4) may be prevented or at least delayed through the detection of discrepancies, since the detection of a discrepancy most probably enhances memory for the sources of the discrepant details. Discrepancy detection, however, is linked to a consistency assumption: Only if I assume that the postevent information is consistent with the witnessed event, I will be surprised by details that do not correspond to it. In contrast, imagine what the case would be if subjects did not assume consistency. The presentation of different response alternatives in a recognition test, for example, clearly implies non-consistency. However, this difference would be no surprise and, therefore, might not lead to a better encoding of the sources. In short, it seems to be a precondition for integration that details representing possible candidates for a certain feature of a witnessed event are not encountered as *discrepant* but as merely *different*. Conversely, with the traditional procedure of deceiving subjects on the presence of discrepant details, integration may actually be prevented. Therefore, a methodological consequence for the study of integration phenomena should be to choose settings where the difference between possible response candidates comes naturally and not as a surprise. If this line of argument holds, it would indeed have enormous consequences, since it implies that a large part of research on eyewitness testimony is simply irrelevant for the integration issue; that is, true integration phenomena had not been studied yet.

Conclusions

Since memory impairment is temporally elusive or small in magnitude, it does, by and large, not seem to be a real threat to eyewitness performance. However, it is certainly of practical importance to know whether and to what extent transformational misinformation effects generalize to real-world situations and what can be done to prevent them. Further, source memory is certainly relevant for an assessment of eyewitness memory in real-world situations since large time intervals between witnessed incident and testimony cannot always be avoided and poor source memory might result in long-time effects of misinformation previously rather neglected in eyewitness research. It would be worthy to establish the impact and generality of such effects. Finally, I suggest that the usual short-term-and-deception designs employed in eyewitness suggestibility research are not suitable for an assessment of integration phenomena. The

latter should be studied in long-term designs without deceiving the subjects on the presence of misinformation.

1) I do not claim to be the first who uses a problem solving metaphor in connection with memory. Norman (1970), at least, used one, though in a more narrow way.

2) This follows from another conversational maxime, the maxime of *relevance* (Grice, 1975): When the Lindsay and Johnson test asks for details which were only present in the narrative, the subjects assume that such details had in fact been presented; otherwise, the response alternative was not relevant.

3) I prefer the term enlightenment over the usual term warning since the former refers especially to the reversal of a prior deception, whereas the concept of warning covers several quite different measures.

4) I want to make clear that I refer to memory impairment throughout this work, following Loftus (1991), as a general notion which covers overwriting or blending as well as inaccessibility of memory traces (trace and retrieval impairment, in Loftus' words). It is true that IMP's memory states reflect the availability of details at the time of test. However, the same practical consequences would arise if the details were not unavailable but merely inaccessible for some reason: the subject could not use this information at the time of test.

5) A word on the power of the chi-square test: Since IMP makes exact quantitative predictions for various assumed degrees of memory impairment (see section A), it is also possible to calculate the exact power of the chi-square test (without source consideration) to detect such impairment if present. The effect size (derived from Cohen, 1988, p. 216) depends on the actual frequencies of simple memory states in the experiment and on the assumed degree of impairment; it equals w = x] * [[f(OD+)]*[f(MD+)/f(OD-)]]1/2[, where x is the amount of impairment and f(OD+) means the frequency of correct recognition of the original details; the other terms are explained likewise. A crucial question is, of course, which amount of impairment should be regarded as a reasonable effect to look for. Belli, Windschitl, McCarthy, and Winfrey (1992, p. 363), for instance, have estimated a 32% impairment in one of his experiments. My own estimations from 10 experiments done with the modified test procedure (including those of Belli et al., 1992; Ceci, Ross, & Toglia, 1987; and Chandler, 1989, 1991) yielded an estimated 40% impairment. To be on the safe side, I have calculated the power for a smaller estimate. In experiment 1, the power (tests with df = 2) to detect 30% impairment was .85 for the discrepancy type slide vs. narrative, and >.99 for both slide vs. tape and narrative vs. tape (linear interpolations from the power table).

6) These compatibility intervals are very much like confidence intervals. However, in the conventional use of confidence intervals, the sample value and the population parameter have the same dimension (e.g., a proportion, a variance, etc.). This is obviously not the case in the present calculations where the population parameter (i.e., the amount of impairment) is more "removed" from the sample value (i.e., the frequency distribution of combined memory states). On the other hand, the logic of estimation is the same (i.e., looking for values of the population parameter which are - at a certain probability level - compatible with the obtained sample value). But it is precisely because of the differences that I have chosen a new terminology.

7) The relation of these data to those in the response patterns is the following: the percentages of simple memory states are entirely deducible from the response patterns. For instance, the 53% memory states I in figure 5 (slide vs. narrative) correspond to the 53% responses in the response C1 in figure 4, and the 3% percent memory states II are the sum of the remaining responses in the same column of that reponse pattern (rounding error!). The frequencies of combined memory states cannot be deduced from the response patterns since these figures reflect the frequency of remembering the original details (and their source)]*conditioned upon* [the frequency of remembering the misleading details (and their source), and vice versa. In contrast, the response patterns give the overall frequencies of remembering the original or misleading details. However, a translation is possible the other way round. For example, the obtained frequencies of the combined memory states 1, 4, and 7 in figure 5 (slide vs. narrative) sum up to the 51% responses in E1.

8) The power for these tests (with df = 3) to detect 30% impairment (see footnote 5) is .52 for the discrepancy type slide vs. narrative, .80 for slide vs. tape, and .74 for narrative vs. tape (linear interpolations from the power table).

9) Four subjects were excluded because I had made a mistake in the conduction of an experimental session. Another subject was excluded because of knowledge about the true purpose of the experiment. Further, one subject in each group was randomly dropped from analysis in order to have equal numbers of subjects in each counterbalancing condition (see design).

10) In fact, this constitutes an "instruction dilemma": Putting too little emphasis on the source runs the risk that

subjects will not follow the instructions, and emphasizing too saliently that they should base their responses solely on the slide will lead them to assume that there might be differences between the two sources, because otherwise the emphasis was irrelevant (or, the maxime of relevance was violated; see Grice, 1975). Thus, it had the same effect as a warning. I wonder why this dilemma has not yet been discussed in the literature, since it is inherent in every test procedure (unless it includes a warning).

11) Actually, the memory states 2 and 4 are not possible to exist at all for traditional designs with two sources of information. Because the subject can find out the second source, they "turn into" memory state 1, that is, the frequency of this memory state would be augmented by the frequencies of the impossible memory states 2 and 4. However, this would not change the predictions, since the single performance within these memory states is assumed to be 100% correct each. Therefore, and for consistency with my previous presentations, I want to retain the memory states 2 and 4 (as "virtual" memory states, so to speak).

12) If the obtained frequencies of memory states in C2 are used as the basis for the simulations, the best fit is obtained for pR1 = 50% and pR2 = 0%.

Part II: Paired-associate learning

II.1. Introduction to part II

It has often been found that people remember a previously encountered piece of information more poorly when they have later been presented similar information. Examples are the well-known phenomenon of retroactive inhibition (RI) in paired-associate learning (see Crowder, 1976, for a review), the misinformation effect in research on eyewitness suggestibility (see part I), and, in part, also the hindsight bias in the domain of judgment and decision making (see Hawkins & Hastie, 1990, for a review). Explanations for these phenomena (I will refer to them jointly as RI effects) have often been sought in terms of memory deficiencies, for example, unlearning, memory impairment, or assimilation of memory traces. While this is certainly true in some form for part of the effects observed, these kinds of explanations have also been questioned in pointing out that other mechanisms than memory deficiencies contribute to the effects as well. For example, a failure to discriminate word-lists has been made responsible for RI effects in paired-associate learning (Abra, 1972). In a related fashion, source monitoring errors were forwarded as an account of the eyewitness misinformation effect (Lindsay & Johnson, 1989). However, it is not clear to what extent the different mechanisms (i.e., memory or source mechanisms) contribute to the overall effects. I believe that theoretical and mathematical modeling can be helpful in clarifying the interrelations of the conflicting theoretical approaches and outline the limitations of their respective contributions.

The purpose of part II is to theoretically specify and empirically demonstrate how failures to discriminate the sources of different pieces of information can yield *test-dependent* RI and PI effects in a paired-associate learning task. This is not to say that I consider this mechanism to be a major constituent of RI and PI effects in general. My emphasis is on demonstrating the possibility of an effect which is *not* brought about by any changes in memory but - putting it very general - the result of asking subjects for one piece of information when they have been presented two, given bad source memory (but not worse experimental than control source memory). This very broad approach also implies that I do not consider this effect to be restricted to paired-associate learning; as already noted, it came to my attention as a result of simulations done with the integrative model of eyewitness performance presented in part I. For purposes of exposition, the model will be introduced once again here, but in a more general fashion that underlines its applicability to both eyewitness suggestibility and paired associate learning. Furthermore, the introduction focuses on those components of the model which are necessary for understanding why and under what circumstances the effect under question should appear. Later, I report two paired-associate learning experiments which try to obtain such an effect with an appropriate design.

II.2. Introduction of the model (Adaption for part II)

The model is based on a central distinction between memory and performance. The latter is seen as the product of a kind of problem-solving activity in which the remembering person tries to find a solution to a given memory task (for instance, "which response term was associated with stimulus term X in list 1?"). Thereby, he or she uses information available in memory. However, performance also depends on external factors which influence (a) the memory search process, that is, what kind of information the subject considers relevant, and (b) the evaluation of available information with respect to its usefulness for solving the memory task. A major external factor, for instance, is information pointing to the consistency or non-consistency of pieces of information from different sources (see below and part I).

Though the formulations given above sound like a retrieval-based explanation - which could probably be handled within conventional memory models - the problem-solving metaphor of remembering also includes processes commonly referred to as response biases (see part I). Therefore, and because the dividing line between "true" retrieval processes and "mere" response biases is thin and probably artificial, I refer to both types of processes jointly as the transformation of memory into performance (this will be detailed below). In principle, then, RI and PI effects can be either consequences of real memory deficiencies, that is, an impaired memory in the experimental condition, or arise from a deteriorated transformation of an intact memory into performance. Part II focuses on the latter.

Memory states

The memory base for performance is thought to consist of certain distinct *memory states* reflecting the availability or non-availability of pieces of information which are relevant for solving the memory task. Most RI or PI memory tasks require subjects to remember a certain piece of information - the target item - and attribute it to a certain source of information; this encompasses, for instance, asking for the list 1 response terms in studies on retroactive inhibition or asking for details from an initially observed event in eyewitness suggestibility research. It is certainly relevant for performance on such tasks whether subjects (a) can remember (or have available; these terms will be used synonymously) the target item and (b) whether they can remember from which source of information it came (for instance, from which list). This allows to distinguish three *simple* memory states for this item: (I) item available, source available; (II) item available, source not available; (III) item not available, availability of source irrelevant. This reflects memory in the control condition of RI or PI designs. However, the crucial feature of all interference designs is that there is always (at least) a second item in the experimental condition - the distractor item - which interferes or not interferes with memory performance for the target item. For this distractor item, three corresponding memory states can be distinguished in the same way. Further, in the experimental condition, where both items

--- 61 ---

are presented, nine *combined* memory states are possible to exist as a result of the combination of 3 * 3 simple memory states for the single items; this is illustrated in figure 1. As a result of, for instance, different encoding across items and subjects, there will be frequency distributions of (simple and combined) memory states; these constitute the memory base which is necessary for running simulations of memory performance under various testing conditions. Figure 1 gives a hypothetical example for such frequency distributions which is derived under the assumption that memory for the target and distractor items is *independent* from each other in the experimental condition, that is, the storage or accessibility of the target item is not touched after the encoding of the distractor item. This means that the frequencies of the combined memory states equal the product of the frequencies of the respective simple memory states for the target item and the distractor item alone (see part I for details).

Distractor item	Target item		
	I: item + source 40%	II: item, no source 20%	III: no item 40%
I: item + source 50%	1 20%	2 10%	3 20%
II: item no source 20%	4 8%	5 4%	6 8%
III: no item 30%	7 12%	8 6%	9 12%

Figure 1. Frequency distributions of memory states (hypothetical example). Roman (arabic) numbers denote simple (combined) memory states.

Transforming the memory base into test performance

The simple logic behind simulating performance on a memory test (i.e., percent correct recognition in the cases studied here) from such a memory base is that the total test performance can be regarded as built up from performance contributions of single memory states; this holds for the control condition as well as for the experimental condition. That is, if we consider how a subject will perform on a given memory test within each (simple or combined) memory state,

we only have to weigh this expected performance with the relative frequency of the respective memory state to have its performance contribution; adding these contributions across all memory states yields the expected total performance. This is expressed in the following formulae:

$$\text{Total control performance} = \sum_{i=I}^{III} f_{ci} * P_{ci} \quad \text{and}$$

$$\text{Total experimental performance} = \sum_{i=1}^{9} f_{ei} * P_{ei},$$

where f_{ci} (f_{ei}) is the relative frequency of a simple (combined) memory state in the control (experimental) condition, p_{ci} (p_{ei}) is the expected single performance within this memory state, and I - III (1 - 9) are the running numbers of the simple (combined) memory states.

Now it is evident that, in order to run a simulation off, two things must be known, (1) the frequency distributions of memory states and (2) the expected performances within memory states. With respect to the former, there are two ways to handle this problem: first, simply assuming them (this is sufficient for the simulation done in the following), and second, estimating them more or less directly from data in an actual experiment (this was done in the experiments reported below). On the other hand, which performances we have to expect within memory states with a certain test procedure must be determined individually for each memory test situation. I want to illustrate this for the test procedure used in experiment 1: In continuity with my previous research (see part I), I used a two-alternative forced-choice recognition test which corresponds to the test procedure used by Loftus, Miller, and Burns (1978) in eyewitness research, a procedure which is (or at least was) widely used in the field. Transferred to paired-associate learning, this procedure would provide the stimulus terms and ask for the associated response terms from a specified list (i.e., list 1 for RI and list 2 for PI), thereby requiring subjects to choose between the list 1 and the list 2 terms. I expect the following performances within simple and combined memory states (see figure 2):

In the experimental condition, performance should be 100% correct within the memory states 1, 4, and 7, where the target item is available and subjects know its source. The same should hold for memory state 2, because even if the target item source is not known from memory, it can be deduced from the available source of the distractor item.[1] In memory state 5,

subjects remember both items but not their sources. Therefore, they have to guess when asked which of them came from the source in question and, consequently, have a 50% chance of being correct. With memory state 9, the subjects remember nothing and must guess anyway.

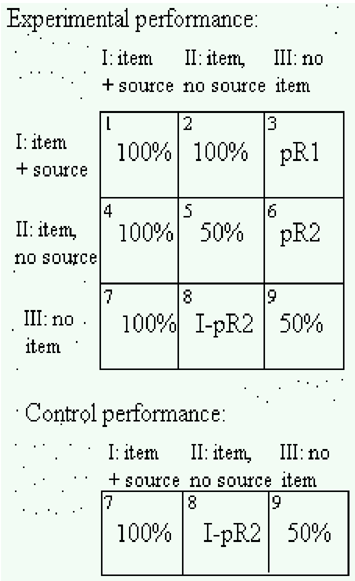


Figure 2. Expected performances within memory states in experiment 1 (two-alternative forced-choice test including the distractor item as alternative).

Memory state 3 depicts a situation where subjects have to rely on additional information in order to solve the task. Their memory holds only the distractor item as well as its source, that is, they know that what they remember is not strictly what the memory test asks for. Thus, their answer critically depends on their knowlegde about the relation of the response terms from the target list and the distractor list. If there were no common response terms (as in traditional AB-AC designs), subjects might infer that the remaining response alternative must be the target item and, consequently, be correct on the test. Alternatively, if there were also identical response terms in both lists, this inference cannot be valid, and so the subjects have, more or less, to guess whether the remembered distractor item is the correct answer. In general, thus, the subjects' behavior within memory state 3 depends on context factors. Therefore, I

have employed a parameter pR1 to denote the probability that the remembered distractor item will be rejected on the test. Simultaneously, this is the probability of being correct on the test (because with the standard test procedure, rejecting the distractor item means choosing the target item).

In the memory states 6 and 8, the subjects remember a certain item they had encountered during the experiment but not its source. That is, compared to memory state 3 where subjects may conceivably reject the distractor item because of their knowlegde that it came from the irrelevant source, here they have no specific reason to be cautious. However, unspecific knowlegde about the presence of different first and second items may nevertheless prevent them from always accepting the remembered item as the answer. Hence, I have employed another rejection probability pR2 for the memory states 6 and 8, though it should be markedly lower than pR1. - While pR2 is assumed to be equal for the memory states 6 and 8, the expected performance is not. This is simply because in memory state 6 the subjects remember the distractor item and are therefore correct when they reject it on the test; in memory state 8, however, they remember the target item and are therefore correct if they do *not* reject it.

Deriving the expected single performances in the control condition is fairly trivial for the simple memory states I and III. In the former, the subjects remember the first item and its source; hence, they will choose it on the test. In the latter, they know nothing and have to guess. With memory state II, the subjects are in the same position as with the memory state 8 in the experimental condition: They remember a certain item, which happens to be the target item, but not its source. Therefore, the same arguments apply as above, and performance is assumed to be identical to memory state 8.

A simulation of RI and PI effects

With the above considerations in mind, we can now simulate performance on the Loftus standard test on the basis of various assumed frequency distributions of memory states. For sake of simplicity, I will assume that there is equal memory for the target items and the distractor items, and I will hold memory for these items at a constant level of 60%, so that the only memory variable is the relative amount of cases where no source is available. Moreover, I assume independence of memory for target and distractor items in the experimental condition, as defined above (cf. figure 1). The rejection probabilities discussed above are set at pR1 = 50% and pR2 = 0%; this (a) imagines a situation where subjects' knowlegde about the source of the remembered second item in memory state 3 does not allow them to make any inferences regarding the question for the first item (i.e., pR1 = 50%)[2]. Further, it is assumed that (b) any

--- 65 ---

"unspecific" rejections in the memory states 6, 8, and II play no major role (i.e., pR2 = 0%). Table 1 lists the performances resulting under these assumptions separately for the control condition and the experimental condition.

Table 1: Simulated Performances for a Two-Alternative Forced-Choice Test Including the Distractor Item as Alternative, as Used in Experiment 1 (in % Correct Recognition; Hypothetical Example; see Text for Details)

Memory states		Performance	
I	II	Control	Experimental
60%	0%	80%	80%
50%	10%	80%	77,5%
40%	20%	80%	74%
30%	30%	80%	69,5%
20%	40%	80%	64%

As one easily sees, the worse memory for the sources, the bigger is the effect. Of course, this holds only ceteris paribus; for example, a higher (lower) rejection rate pR1 would diminish (enlarge) the effect. Moreover, the effect is - trivially - diminished (enlarged) when memory for the target items is better (worse) than the assumed 60%. Anyway, the figures presented in table 1 give a good impression of the magnitude of the effects to be expected: They are not dramatic but nevertheless big enough to raise interest since they may at least sometimes contribute to overall RI and PI effects found in studies on paired-associate learning.

Since the language of target and distractor items does not depend on any temporal sequence of items, the simulated effects could be RI as well as PI effects. However, RI and PI effects are identical in size only if there is identical memory (i.e., identical frequency distributions of simple memory states) for target and distractor items. In general, given better first list than second list memory, RI (with first list items being the targets) should be be smaller than PI (with second list items being the targets), and vice versa.

II.3. Experiment 1

Experiment 1 was designed as a first attempt to find empirical support for these simulated effects. The logic was to run a simulation like that presented above, but on the basis of *empirical* frequency distributions of memory states, and compare its results to empirically obtained performance on the Loftus standard test. Of course, this requires some way of getting access

--- 66 ---

to the (frequency distributions of) memory states; to this end, I employed a separate memory test (henceforth called memory test A) which asks simultaneously for target as well as distractor items (i.e., response terms) and their sources (i.e., lists) and allows for a one-to-one mapping of answers to memory states, in the control condition as well as in the experimental condition.

There are some unusual features of experiment 1. First, I used three sources of information, that is, three successively presented lists of word-pairs. This is because memory test A in the form I used it since (see part I) requires to have more than two sources; the test asks subjects (in the experimental condition) to attribute two conflicting items to their respective sources; this decision would be trivial if there were but two sources and the subjects know the source of at least one item; consequently, source memory could not be reasonably assessed. However, as in traditional RI and PI studies, there were only two conflicting responses associated with the same stimulus. Second, experiment 1 was designed for an assessment of RI effects as well as PI effects, corresponding to the twofold simulations of the model. As a consequence, there were also control conditions with respect to an assessment of PI; more precisely, there was a control condition for each single list. Correspondingly, memory test B consisted of three subtests B_1, B_2, and B_3 which tested memory for the response terms from the respective lists. Finally, another condition was added whose only purpose was to assure that performance in the combined memory state 3 (pR1) was somewhere near the 50% which were assumed in the simulations. In this condition, two *consistent* response terms were paired with the same stimulus term. This should prevent subjects from inferring the correct answer in the two-alternative forced-choice test from knowing the false answer, which was possible if the target response term was always different from the known distractor response term.[3]

Method

Subjects. 28 students and staff members from various faculties of the University of Constance took part in the experiment (19 male and 9 female, aged from 20 to 35 years; *Md* = 26 years); most of them were paid for participation.

Design. Three lists of word-pairs were used in the following way: A given stimulus or A term could be paired with a response term from list 1 (X terms), from list 2 (Y terms), or from list 3 (Z terms). However, no more than two response terms were paired with the same stimulus term across lists. This allowed to have three experimental conditions, corresponding to the

possible combinations of two lists each: list 1 vs. list 2 (AX-AY), list 1 vs. list 3 (AX-AZ), and list 2 vs. list 3 (AY-AZ). Three control conditions consisted of word-pairs presented in one list with no interfering pairs in other lists: AX (list 1), AY (list 2), and AZ (list 3). Finally, a seventh condition was run in which subjects received consistent word-pairs in each of two list-combinations; this condition was further subdivided into three "subconditions": list 1 vs. list 2 (AX-AX), list 1 vs. list 3 (AZ-AZ), and list 2 vs. list 3 (AY-AY). All of these seven conditions were run within subjects, with different pools of word-pairs assigned to conditions across subjects according to a latin square. That is, every subject was run in every condition, but with different groups of word-pairs, and every group of word-pairs was used in every condition, but with different subjects. Between subjects, the memory (sub-)tests A, B_1, B_2, and B_3 were presented in four different temporal orders (A-B_1-B_2-B_3; A-B_3-B_2-B_1; B_1-B_2-B_3-A; B_3-B_2-B_1-A) to control for order effects.

Materials. Each pool of word-pairs consisted of nine stimulus terms (A terms) and nine times three response terms (X, Y, and Z terms) which were used in the three lists, respectively. The stimulus terms were more or less explicit category terms (e.g., color, bed, smoking, planet, fruit, prophet), and the response terms were members of these categories or objects which could at least be associated with the respective terms (e.g., turqoise, lilac, pink; pillow, sheet, coverlet; ash-tray, cigar, lighter; etc.). Based on pre-tests and occasionally on face-validity, more untypical category members or associated objects were chosen, in order to preclude, as much as possible, hits due to typicality-based guessing on the memory tests. Furthermore, the word pools were roughly matched for memorability on the basis of the pre-tests. - Combining the groups of word-pairs in the respective counterbalancing conditions yielded seven times three lists of 33 word-pairs, each printed on a single sheet. The order of the word-pairs was randomly determined within each list.

Procedure. Subjects were tested individually or in small groups up to four; they were assigned to the counterbalancing conditions according to a predetermined schedule, with the restriction that within one session only one temporal order of memory tests could be realized. At the beginning of the session, subjects were informed that they would have to study three lists of word-pairs for three minutes each, with breaks of one minute between the lists, and that they had to remember these lists after an hour. The precise nature of the memory test was not revealed to the subjects. In group sessions, subjects were informed that each of them would receive individual lists. When studying the lists, subjects were reminded when appropriate that one minute (respectively, 20 sec) was left for study, in order to ensure that the lists were worked through completely. After the study phase, subjects were required to return in exactly one hour; subjects in group sessions were asked not to talk with each other about the lists during this time. All subjects except two managed to return within an hour plus or minus 5 min; these subjects returned after 70 min and 75 min, respectively. The subjects were announced several memory tests in which the stimulus term of each word-pair would be presented to them; their

task would be to provide the appropriate response terms. Additionally, they were reminded that each given stimulus term could have appeared once or twice in the lists; if presented twice, the response terms associated with them could have been different or identical; the main reason for this briefing was to prevent subjects from suspecting that there might have been three or zero response terms associated with a stimulus term; this would have raised problems for the analysis of memory test A results. Then subjects began with the first memory test; when finished, each memory (sub-)test was removed before subjects began with the next. All tests were self-paced; the subjects took between 20 min and 40 min approximately to finish all of the tests.

Memory tests. Memory test A provided all stimulus terms that were presented in the lists. For each stimulus term, subjects had to choose among seven (in rare cases six) substantial response alternatives and a further, "don't know" alternative. They were instructed to (a) underline the one or the two reponse terms that had been paired with the stimulus term in the study phase (or "don't know") and (b) write behind the underlined response term the number of the list where it had appeared (or "?" if they did not remember); (c) if they thought that a certain response term had appeared twice, subjects were instructed to write both lists behind this term. Subjects were told that they had not to be sure to indicate a certain response alternative or list; however, if they had the feeling that their answer would be a complete guess, they were instructed to indicate "don't know" or "?". - In contrast to memory test A, the memory subtests B_1, B_2, and B_3 asked for the response terms from only one list each. These tests were printed on separate sheets and entitled "memory test B - list 1" ("2", "3", respectively). All stimulus terms from the respective list were provided, and subjects had to decide which one of two response alternatives had been associated with the respective stimulus term in this list. For stimulus terms used in the experimental conditions, the response alternatives were those response terms that had in fact been used in the study phase (i.e., the target item and the distractor item). For stimulus terms used in the control conditions or in the consistent condition, one of the response alternatives was the correct one (e.g., X), and the other was chosen from the remaining response terms used in other conditions (i.e., Y or Z); thereby, each of these remaining response terms was used equally often. Subjects were instructed that they definitely had to choose one of the alternatives even if they thought that they were merely guessing.

Memory state analysis. Simple and combined memory states were identified with memory test A as follows: each properly underlined response term was taken as evidence that the respective response was available to the subject; conversely, false underlinings and "don't know" responses indicated the non-availability of the response. The same logic holds for source memory: correct list indications were taken as evidence for the availability of the source, and false, "?", or (rare) missing list indications were taken as evidence for the non-availability of sources. All underlined responses were counted fully except when two sources were noted behind this term; then the response was splitted into two half responses. For example, if a subject had underlined the correct response in a control condition but indicated two sources, the correct and a false one, this was counted as half a simple memory state I and half a simple memory state II. The same logic holds for the experimental conditions.

Results and discussion

Order effects. There were no significant main effects of the temporal order of memory tests (most *F*'s < 1), neither with respect to tests A and B nor with respect to the order of subtests within test B, as revealed by separate ANOVAs with individual amount of RI or PI in test A or B as dependent variable (absolute difference between numbers of correctly recognized response terms in the respective control and experimental conditions); test (subtest) order was treated as fixed (random) effect. However, there was one significant interaction between test order and subtest order for the PI results in test B ($p < .05$). Since this was the only significant effect of test orders on performance and, moreover, difficult to interprete, I decided to ignore it. - These results may be surprising since one could expect (a) some learning from one test to the next and (b) that the presence of response alternatives in test B influences subjects' choices in test A if test B preceded test A. Maybe such effects did not matter because (a) the overall memory level was fairly high so that further improvement was hardly possible and (b) the tests were too long to permit such kinds of inferences. Anyway, order effects are not considered any further, and all results reported below are collapsed across orders.

Pooling. Furthermore, to ensure economy of reporting and to raise power for statistical analyses, all individual comparisons of control and experimental conditions were grouped together. Remember that, with respect to RI, there were three possible comparisons of performance: AX vs. AX-AY, AX vs. AX-AZ, and AY vs. AY-AZ. Thus, RI results involve the comparison of (2 * AX + AY)/3 vs. (AX-AY + AX-AZ + AY-AZ)/3; the same logic holds with respect to PI. Taken together, each experimental mean presented in this section is based on 3 (comparisons) * 9 (word-pairs per comparison) * 28 (subjects) = 756 observations, whereas each control mean is based on 504 observations (only two comparisons, though the results of one were counted twice). - The results of the consistent condition are of minor interest and not reported here.

RI and PI. Control and experimental performance in memory tests A[4] and B are given in table 2. One easily sees that the expected and simulated effects did in fact show up: there is significant RI and PI in memory test B ($t = 4.77$ and $t = 2.58$, respectively), while there is no significant amount of PI ($t = 1.37$) and even a reversed RI effect ($t = -2.21$) in memory test A[5],

as revealed by *t*-tests for paired observations with the number of correct responses per subject and condition as the dependent variable. However, these are merely the "surface" results; there are more questions to be asked. First, how to explain the somewhat counterintuitive reversed RI effect in memory test A? There is a post-hoc explanation[6] which I find very plausible: When subjects in the experimental condition, during reading the second and third list, encountered a conflicting second response term associated with a given stimulus term, they were surprised and tried to call back into their minds what the first response term had been, thereby giving it an additional rehearsal. Moreover, the time needed for these rehearsals reduced the actual study time for the second (or, third) list, what nicely accounts for the PI effect in memory test A which is similar in magnitude but in the opposite direction.

Table 2: Mean Percent Correct Recognition in Experiment 1

	Memory test		
	A	B	B simulated
RI			
Control	81	91	90
Experimental	85*	83***	84
PI			
Control	76	84	88
Experimental	73	76*	77

* = p < .05, *** = p < .001; t tests, df = 27, two-tailed.

[Table 2]

Simulation of test B results. The most important question, of course, is how the performance differences between tests A and B emerged. Did they originate the way I have assumed in the above simulations? If so, then doing the same simulations (i.e., with the same assumed performances within memory states) on the basis of the empirical frequency distributions, as recorded with memory test A (given in figure 3), should lead to memory test B performances that come close to the actual ones. Indeed, this was the case, as can be seen in the right column of table 2. The fit is virtually perfect except for the simulated control PI performance which differs from the actual value by four percent points. However, this does not necessarily mean that all of the model's assumptions are confirmed, since one could imagine different - though less plausible - assumed performances within memory states which would yield similar simulations.

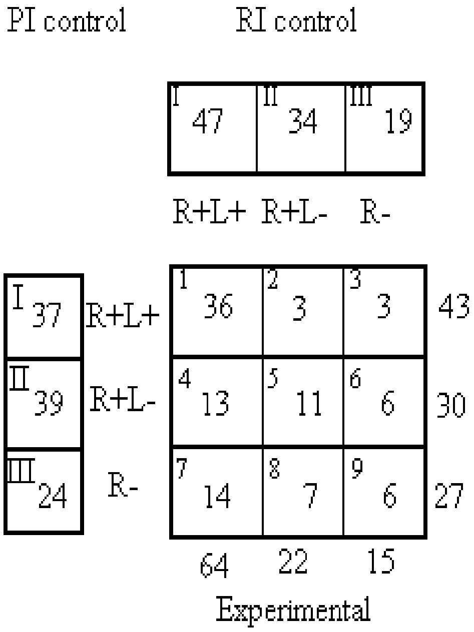


Figure 3. Relative frequencies of memory states in experiment 1. Numbers are percentages. R+L+(-) means the ability to remember a response term and (but not) its source. R- means that the response term is not available.

Empirical performance within memory states. Therefore, a more direct test of the assumed performances within memory states is wanted. Fortunately, such a test is possible since each subject was administered both memory tests A and B. We simply have to identify a subject's memory state on each test A item and register the subject's choice on the same item in test B. Pooling across items and subjects then allows to determine the proportion of correct test B

answers per (simple or combined) memory state as identified with test A. These figures are precisely the empirical performances within memory states; they are given in table 3 and contrasted with the performances which had been theoretically set in the simulations. Though there is certainly a qualitative fit, some deviations nevertheless require comment. I see two types of deviations, ones with methodological significance and ones with substantial significance. Of the first type are the deviations in the simple memory states I and III and in the combined memory states 1, 4, 5, 7, and 9: they reflect failures of memory test A to adequately identify memory states. That is, among the responses classified as memory states 1, for instance, could have been a few unsure memories or hits due to chance; it is conceivable that these had led to suboptimal performance (i.e., less than 100% correct) on memory test B. Similarly, the above-chance performance in memory states III, 5 and 9 indicates that subjects sometimes knew more than would have justified their classification into these memory states.

Of more importance are the deviations I have called substantial; this refers (a) to within-memory-state performances that are thought to depend on the parameter pR2; this covers the combined memory states 6 and 8 as well as the simple memory state II. Performance within the former is essentially homologous, as hypothesized, though with pR2 at an unexpectedly high 25% on average (this homogeneity leads me to ignore the small statistical basis for each single proportion). This could mean that subjects were reluctant to accept any one remembered response without knowing its source. However, this would not explain why performance is clearly better in the simple memory state II (though one might see the 90% in the PI control condition as a trend in the right direction, which, besides, accounts for the misfit of the PI control simulation in table 2). My post-hoc explanation for this difference is that in the experimental condition, subjects sometimes knew that a second reponse term had been presented though they could neither remember what it was nor where it came from (this mere knowing about a second response is revealed by their giving additional "don't know" answers); in the control condition, this was not possible, since by definition there was no second reponse term. This knowledge about the presence of a second reponse term may have reminded the subjects in the experimental condition that the remembered response term might not be the one presented in the list under question and, in a number of cases, they decided that it must have come from another list; consequently, they opted for the other response alternative.

(b) The second major deviation is the lowered performance in the combined memory state 2, which came to me as something of a surprise. Here, subjects could in principle have reached perfect performance by inference: They knew that they had not been presented more than two response terms, and knew about the source of one of them; hence, they only had to respond with the other when asked for the term from one of the other lists. But perhaps the memory tests were too time-consuming to perform such kinds of inferences, and so subjects might have guessed if they could not immediately reach at a decision. - Finally, there are not only deviations. It should be noted that there is nearly perfect correspondence of assumed and obtained

performance in memory state 3, where performance depends on the parameter pR1. The performance around 50% indicates that the subjects had not been too sensitive to the actual frequencies of consistent and different response terms, though the trend is in the right direction.

Table 3: Performances within Memory States (Percent Correct Recognition) in Experiment 1

Memory state	Performance Assumed	Obtained RI	PI
Control			
I	100	98 (96-99)[b]	96 (92-98)
II	100[a]	96 (92-98)	90 (85-94)
III	50	66 (57-74)	55 (47-64)
Experimental			
1	100	97 (94-98)	96 (92-97)
2	100	76 (57-89)	79 (70-86)
3	50	58 (38-76)	56 (47-65)
4	100	97 (92-99)	92 (75-98)
5	50	66 (55-75)	63 (52-72)
6	0	20 (11-34)	30 (20-43)
7	100	95 (89-98)	96 (79-99)
8	100	78 (65-87)	80 (66-89)
9	50	53 (39-67)	56 (41-70)

[a]Figures in italics are assumed parameter values

[b]Figures in parentheses are the 95% confidence intervals for the proportions (calculated after Hays, 1973, p. 379)

--- 74 ---

To sum up the results and discussion of experiment 1: the major result was the lack of RI and PI in memory test A and the presence of RI and PI in memory test B. That is, the effects predicted from a priori simulations were qualitatively confirmed. Further, a posteriori simulations based on the empirical frequency distributions of memory states yielded results that closely matched the actual performances in memory test B, suggesting that the effects could in fact have come about in the way described by the model. Closer inspection of the empirical within-memory-state performances essentially supports the model assumptions except for minor deviations which reflect, on the one hand, occasional unreliability of the memory state classification procedure, and, on the other hand, additional influences which, however, would inflate the model if taken into account. Positively, the empirical within-memory-state performances allow for a precise decomposition of the RI and PI effects in memory test B into their major constituents: Table 3 reveals that the "losses" in the experimental condition compared to the control condition are due to the memory states 2, 5, 6, and 8. All of these, however, are memory states with bad source memory. This should be the most obvious confirmation of the conjecture made at the beginning of this article, namely, that there are test-dependent RI and PI effects due to bad source memory (I want to emphasize again that this does not mean worse experimental than control source memory; figure 3 shows that even the reverse is the case!).

II.4. Experiment 2

However, the opening conjecture went even further. I argued that there are RI and PI effects that result from asking for one item from one source when two were presented, given bad source memory. This latter claim cannot be ultimately decided upon on the basis of the experiment 1 results, since the memory tests A and B used in experiment 1 differed in two respects: First, memory test A asked for both response terms while memory test B asked only for the response term from a given list. Second, test A was a multiple-choice test with an additional "don't know" option, whereas test B forced subjects to choose between two alternatives. Experiment 2 allows for disentangling these two variables by offering the subjects the same response alternatives in test B as in test A; that is, the tests are otherwise identical but differ only with respect to the asked-for information: Test A asks for both response terms like in experiment 1, and test B asks only for the response term from a given list. Thus, test A may be called a *global* test procedure and test B a *local* one. If the forced choice procedure used as test B was responsible for the RI and PI effects in experiment 1, then there should be no such effects in experiment 2 (even with bad source memory given). Conversely, if the global/local distinction is crucial, then we must expect to find RI and PI in memory test B (given bad source memory).

--- 75 ---

Experimental performance:

	I: item + source	II: item, no source	III: no item
I: item + source	1 100%	2 100%	3 0%
II: item, no source	4 100%	5 (1-pR3) × 50%	6 100%
III: no item	7 100%	8 1-pR4	9 0%

Control performance:

I: item + source	II: item, no source	III: no item
7 100%	8 1-pR4	9 50%

Figure 4. Expected performances within memory states in experiment 2 (multiple choice test with "don't know option and including the distractor item as alternative).

But there is another way to approach the problem. Like in experiment 1, we can simulate the effects a priori, on the basis of theoretical assumptions concerning the within-memory-state performances with the new test B procedure. I expect the following performances (see figure 4): Within the combined memory states 1, 4, and 7, performance should be perfect since the subjects know both the correct response term and its source. Within memory states 3, 6, and 9, there should be no correct answers in essence; it is conceivable that some subjects guess and make a few hits, but mainly they should choose the "don't know" option. Memory state 2 is treated in the same way as in experiment 1 for the Loftus standard test, that is, perfect performance based on inferences is assumed. Finally, the performances within memory states 5 and 8 depend on parameters. These could be interpreted as "conservatism parameters"; they reflect subjects' tendency to avoid a decision when remembering one or two response terms but not being sure about the source(s) (I have chosen two different parameters since it is not clear whether the amount of "conservatism" is equal in both memory states). Then, when asked for a particular response term from a given list, they simply might escape a decision by choosing the

--- 76 ---

"don't know" option. Of course, this has the consequence that the number of correct answers is lowered by this portion. In memory state 5, performance is further multiplied by 50% since in only half of the cases the subject would choose the correct response term if not being conservative. The assumed performances in the control conditions require no special justification except the indication that (like in experiment 1) the subjects' situation in memory state II is thought to parallel that in the combined memory state 8.

Simulated RI and PI effects based on these assumptions and on the same frequency distributions of memory states as in the simulations preceding experiment 1 are presented in table 4. The parameters were set at pR3 = 50% and pR4 = 0%, without any sophisticated theoretical argument but following the simple deliberation that subjects should be less motivated to avoid a decision when they have no conflicting response terms in memory. The resulting RI and PI effects are (ceteris paribus) even less dramatic as in the simulation preceding experiment 1 (interestingly, manipulations of pR3 and pR4 would have opposite effects: Increasing pR3 increases RI and PI, but increasing pR4 decreases RI and PI; this is because it also influences control performance). Experiment 2 will show whether such effects can be obtained.

Table 4: Simulated Performances for a Multiple Choice Test with "Don't know" Option and Including the Distractor Item as Alternative, as Used in Experiment 2 (in % Correct Recognition; Hypothetical Example; see Text for Details)

Memory states		Performance	
I	II	Control	Experimental
60%	0%	60%	60%
50%	10%	60%	59,25%
40%	20%	60%	57%
30%	30%	60%	53,25%
20%	40%	60%	48%

Method

Subjects. 28 students and staff members from various faculties of the University of Constance served as subjects in experiment 2 (18 male and 10 female, aged from 22 to 37 years; *Md* = 25.5 years); most of them were paid for participation.

--- 77 ---

Design, materials, procedure and memory test A were exactly the same as in experiment 1 (except that this time all subjects returned within 60 min); even the consistent condition was retained though it had no function in the context of experiment 2. The new memory test B was identical to the procedure used in experiment 1 except for the response alternatives. That is, instead of forcing the subjects to choose between two alternatives, they were offered the same response alternatives as in memory test A, including the "don't know" option (see the description of memory test A for experiment 1).

Results and discussion

Order effects and pooling. As in experiment 1, there were no main effects of (sub-)test order. Moreover, there were no interactions of test and subtest order in experiment 2. Therefore, the following results are collapsed across (sub-)test order. Likewise, the RI and PI results are pooled across individual list comparisons.

RI and PI. Table 5 shows control and experimental performance in memory tests A and B (corresponding to table 2 for experiment 1). As already hypothesized from the simulations, the effects are smaller as in experiment 1 but nevertheless present. The dissociation of memory test A and memory test B results is not as clear as in experiment 1[7], even though there were significant RI and PI effects in test B ($t = 2.66$ and $t = 2.74$, respectively); this is mainly due to the test A results: There is no reversed RI effect like in experiment 1 ($t = 0.79$), and there is a larger though nonsignificant ($t = 1.90$, $p = .07$) PI effect in the present experiment. This difference is surprising since the procedure and memory test A were completely identical in both experiments.

I have a post-hoc explanation for this difference which focuses on the lower overall level of performance in experiment 2 (since the procedure was identical, this lower level can only be attributed to the different samples) and on a similar effect found in part I (experiment 2). The main argument is that subjects who were informed about the number of presented response terms (i.e., one or two) in general but not on each individual test item adjust their answers across items to a mean number of one and a half answers. This leads (a) to additional "don't know" answers for control items and (b) to less motivated memory search after further responses for experimental items, if subjects had already remembered one, thus yielding poorer experimental performance. Now it is clear that such a mechanism must lead to greater adjustment if there is more uncertainty to begin with, that is, poorer memory. This would explain why this effect played only a minor role or none at all in experiment 1 where memory was on a very high overall level; in contrast, overall memory in experiment 2 was sufficiently poor to allow for such an adjustment effect. Taken alone, however, it could not wholly explain the memory

--- 78 ---

test A results, since it would suggest symmetrical RI and PI effects. But this can be easily accounted for if we assume that the adjustment effect is superposed by the differential rehearsal effect for the first of two corresponding response terms already discussed for experiment 1: adjustment and differential rehearsal would then cancel each other out for RI and add together for PI.

Table 5: Mean Percent Correct Recognition in Experiment 2

	Memory test		
	A	B	B simulated
RI			
Control	72	67	72 (69)[a]
Experimental	70	60*	66 (65)
PI			
Control	61	57	61 (58)
Experimental	55	47*	51 (50)

* = $p < .02$; t-tests, df = 27, two-tailed.

[a]Figures in parentheses result if pR4 is set at 10%

Simulation of test B results and empirical performance within memory states. Table 5 shows that the simulations on the basis of the empirical frequency distributions of memory states in experiment 2 (see figure 5) are not as satisfactory as in experiment 1, even if the parameter pR4 is somewhat adapted. Some reasons for this misfit are revealed when looking at the empirical performances within memory states in table 6. The most obvious deviation from the assumed values occurs within the combined memory state 8, where the subjects gave the correct answer in only slightly more than half of the cases though they could remember it; this holds for both RI and PI. Closer inspection of the remaining answers shows that here subjects mainly chose the "don't know" option instead of guessing. That is, they were fairly conservative. Interestingly, the amount of conservatism was lower in the simple memory state II which was thought to parallel the combined memory state 8. The same discrepancy showed up in experiment 1; therefore, the same post-hoc explanation is invoked again: in the experimental condition, subjects often knew that they had been presented a second response term, and though

--- 79 ---

they could not remember it, it sometimes prevented them from accepting the other one they could remember.

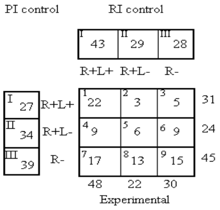


Figure 5. Relative frequencies of memory states in experiment 2. Numbers are percentages. R+L+(-) means the ability to remember a response term and (but not) its source. R- means that the response term is not available.

The other assumed parameter value, pR3, also failed to meet with the empirical performance in the combined memory state 5. Keeping the respective discussion for experiment 1 in mind, however, this appears to be partially the consequence of the minor reliability of the classification procedure. Given that sometimes the temporal sequence of the response terms had remained intact in subjects' memories (as reflected in the above-chance performance in experiment 1), it makes no wonder that they achieved a relatively high level in spite of often being conservative. However, even from this corrected level, an assumption of pR3 = 50% is not tenable. In fact, subjects chose the "don't know" option only 12% of the time in this memory state; this holds for RI as well as for PI. Thus, contrary to my initial assumptions, subjects were more conservative when they could remember only one response term but probably knew of the presence of a second than when they could remember both response terms but not their sources.

With respect to the other within-memory-state performances, the respective remarks in the discussion of experiment 1 hold respectively and need not be repeated here (with the trivial exception that performance in memory states III, 3, 6, and 9 is now close to zero, as expected, rather than around 50% as in experiment 1). The general conclusion from experiment 1 also applies to experiment 2, namely, that the RI and PI effects largely stem from memory states with bad source memory (in this case, the combined memory states 2, 5, and 8).

Table 6: Performances within Memory States (in Percent Correct) in Experiment 2

Memory state	Performance Assumed	Obtained RI	Obtained PI
Control			
I	100	99 (97-100)[b]	98 (93-99)
II	*100*[a]	81 (74-87)	85 (79-89)
III	0	5 (3-10)	4 (2-7)
Experimental			
1	100	96 (92-98)	95 (90-97)
2	100	72 (52-86)	89 (79-94)
3	0	5 (1-16)	6 (3-12)

4	100	97 (90-99)	96 (80-99)
5	*25*	40 (26-54)	51 (37-65)
6	0	6 (2-14)	9 (5-16)
7	100	91 (85-95)	78 (63-88)
8	*100*	61 (51-70)	54 (42-65)
9	0	7 (4-13)	4 (2-10)

[a]Figures in italics are assumed *parameter* values

[b]Figures in parentheses are the 95% confidence intervals for the proportions (calculated after Hays, 1973, p. 379)

--- 81 ---

II.5. General discussion of part II

The results of the present experiments parallel long-known phenomena in the study of memory, especially in the field of verbal learning. Heine (1914) had already found that there is no RI with a recognition test procedure (though Schmidt, 1986, identified a small amount of RI in her data when reanalyzing them with signal detection measures). Similar results were obtained in the 1960s in research on paired associate learning (e.g., Postman & Stark, 1969). These results are in accordance with the small (memory test B) or non-existent (memory test A) effects obtained in experiments 1 and 2. On the other hand, with respect to the difference between the memory tests A and B, it is known that asking for both responses terms, as with the MMFR procedure (modified modified free recall, Barnes & Underwood, 1959), reduces RI in recall tasks. It was assumed that this is because the MMFR procedure eliminates response competition. Also, it is long known that failures to discriminate word-lists lead to the intrusion of irrelevant response terms when subjects are asked for the response terms from a particular list (Abra, 1972).

What, then, is new with the present research? First, to my knowledge, this is the only study to directly investigate the impact of a global (both responses) vs. local (only responses from a given list) test procedure on *recognition* memory. One might speculate whether divergent results in the literature (i.e., no RI effects vs. small RI effects on recognition memory) are related to this distinction. Second, the fine-grained analyses of performance within memory states allow for a precise location of the origin of RI and PI effects. For example, it can be argued that the effects found with memory test B are *not* due to response competition. This can be concluded from the fact that performance in the combined memory state 1 (both response terms as well as their sources available) was virtually perfect in both experiments, that is, there was neither RI nor PI in this memory state. The overall effects in memory test must therefore be attributed to other memory states, in this case, such with bad source memory. However, an analysis of performance within these memory states shows that subjects' failure to discriminate between lists has no uniform consequences. For example, a good deal of the overall RI and PI effects stemmed from subjects' knowledge in the experimental condition that a second response term had been presented to them even if they could not remember it; this knowlegde caused them to be more doubtful about the remembered response term (and to sometimes withdraw from it) than in the control condition where there was no presence of a second response term. That is, we have a differential effect of bad source memory which depends on additional knowledge about the number of presented response terms.

Third, a striking result of both experiments 1 and 2 is that the same psychological mechanisms seem to be responsible for the observed RI and PI effects, as revealed by the empirical performances within memory states which closely parallel each other for RI and PI. Thus, subjects' behavior in memory test B solely depended on their memory for target and distractor

--- 82 ---

items (and additional knowledge) but not on the temporal sequence of these. Of course, this is not to say that this is always the case. It is well documented that there are short-time RI effects (up to some twenty minutes) which were interpreted as reflecting the inertia of a response-set suppression mechanism (Postman, Stark, & Fraser, 1968) and which, on logical grounds, have no PI counterparts. However, if we leave this temporal region (as I have done in the present experiments by choosing a one hour retention interval), the temporal sequence of items per se becomes less and less important compared to the mere fact that two items have been presented which are potential candidates for an asked-for response.

On the whole, it seems that we have to do with two kinds of memory processes. One kind refers to changes in the availability of one sample of items (a list of word-pairs, for example, or an initially witnessed event in eyewitness research) as the result of encoding a second sample of items (a distractor list or a verbal narrative of the event). Such processes have been variously termed the unlearning of responses (e.g., Postman, 1961), response set suppression, blocking (e.g., Niemi, 1979), recoding (Tulving, 1983), memory impairment (Loftus, 1991), and so on. Some of these processes seem to be temporally restricted and reversible. Another kind of memory processes has to do with the transformation of available responses into performance, as already described in the introduction of the model at the beginning of this paper. Here we find a more problem-solving-like kind of activity, that is, more or less deliberate decisions whether to accept a remembered item as answer, depending on additional knowlegde about the number of conflicting or consistent items, for instance. It is these processes I have focused on in part II. Maybe this distinction is equivalent to Tulving's (1983, pp. 175ff) division of retrieval processes into two subprocesses termed *ecphory* and *conversion*: While the former describes the interaction of a retrieval cue with a memory trace, often accompanied by a conscious recollective experience, the latter refers to the transition of the ecphoric information into an overt response, its "use in ongoing behavior".

Given the validity of this distinction, my message is that research on RI and PI (as well as its contemporary counterpart in research on eyewitness memory, for instance) has to date mainly been interested in ecphory processes at the expense of conversion processes. I believe that this has at least two disadvantages: First, even if one is primarily interested in ecphory, one has to face the problem that each observed response is touched not only by ecphoric processes but by conversion processes as well. "Decontaminating" the "real" memory from these conversion processes is easier when we have clear impressions of the latter and do not try to control for various "response biases" in an ad hoc fashion. The model as presented in part II tries to give a systematic account of conversion processes in two not uncommon (at least within memory research) retrieval situations. Thus, though it is not primarily concerned with ecphory and explanations of RI and PI effects that focus on memory deficiencies, this topic is adressed implicitly by pointing out what is not due to memory deficiencies, thus limiting the power of this kind of explanation from without. Second, conversion processes are also interesting in their

own right, especially with respect to applied issues. In daily life, we often have to remember events or pieces of information that were encountered and encoded before time spans that vastly exceed the retention intervals which are commonly used in laboratory experiments. Given that memory for such information and especially for the source(s) of this information gets more and more unsure over time, it seems plausible that there is room for deteriorations in the conversion of such unsure memories into overt answers. I have discussed such conversion processes in connection with eyewitness testimony in part I, under the heading of transformational misinformation effects.

Particularly, remember that the research in part II was partially motivated by an interest in potential long-time effects of misleading information on eyewitness reports which resulted from some observations and speculations reported in section I.6 of part I. It was held that there may be test-dependent effects of post-event information that vary with the quality of source memory like those now found in part II (even if transformational effects due to the subjects' consistency assumption and memory impairment play no role). I do not doubt that these effects would also show up with materials used in eyewitness experiments and in real eyewitness testimony; the reason for this belief is that the effects have shown to depend on two features that are not specific to paired-associate learning: loss of source memory and a memory test that asks for items from but one of several sources of information.

As a first step towards verification, Blank, Wienholt, and Sporer (1995) investigated the impact of "misinformation" presented as response alternatives in an initial memory test on performance on a subsequent test. Such a presentation is a natural way to introduce differing information without allowing for a consistency assumption. Subjects were presented four colour slides (two of them were those used in part I) and received an initial two-alternative forced-choice memory test for 20 items from the slides. After three weeks, they were presented with a four-alternative multiple choice test with an additional "don't know" option and asked what they remembered from the slides. For subjects in the experimental condition, this test contained both response alternatives from the initial test as choice options. Furthermore, there were two control conditions. For 10 items in the zero condition, the subjects had received no initial memory test. Thus, they did not receive the "misinformation". But also, they had no chance to rehearse the original information. Therefore, in a "modified" condition, the subjects received the same initial test but received no "misinformation" in the sense that they had no opportunity to choose it on the test; the proceeding in this condition mirrors the modified test procedure of McCloskey and Zaragoza (1985) except for offering more alternatives and a "don't know" option. Altogether, 30 items were used (whose assignment to conditions was counterbalanced across subjects), and 44 subjects participated. Now, if there are long-time effects of the false alternatives from the initial test - like the effects demonstrated in experiments 1 and 2 - then performance in the experimental condition should be lower than in the "modified" condition (maybe even lower than in the zero condition, if there is not too much improvement through

rehearsal). Results showed that second-test performance in the zero condition was worst (33% correct), indicating that the opportunity to rehearse the original information in the initial test clearly improved performance. Experimental performance, however, was significantly ($p < .05$) lower than "modified" performance (72% vs. 78% correct). More interestingly from the present point of view, experimental performance in a third memory test (administered immediately after the second) which asked for both the original information and the false alternative from the initial test (like the memory tests A in experiments 1 and 2) helped to raise memory for the original details to 85% correct. Analysis of source attributions in the third test revealed that the difference in experimental performance from the second to the third test was essentially due to source confusions. Thus, test-dependent RI effects resulting from poor source memory are also obtained with materials and procedures that are more close to real eyewitness testimony.

One may speculate whether the impact of "misinformation" (or, perhaps more correctly, *rival information* - that is, information which is a possible but false candidate for a test answer without being misleadingly suggested as the correct answer) goes beyond being presented within the same context as the original information (i.e., as two response alternatives in the initial test). For example, would rival information presented in an unrelated story also affect performance after some longer time interval? If yes, this would be bad perspectives for real testimony, since it is difficult to imagine that anyone could escape any information after having witnessed a crucial event. Performance would then depend on the probability that rival information is available in the time after the event.

In any case, given the long time intervals that often lie between the encoding of relevant information and a testimony, and that source memory diminishes rapidly with time, there might well be serious damages of performance if the interrogation process focuses exclusively on information from one source, that is, the original event, a proceeding which corresponds more or less to the local memory tests B used in experiments 1 and 2. That is, such a procedure will invite source attribution errors and, furthermore, may create an illusory validity of the obtained information (since the uncertainty of source memory is not recognized). We might contemplate whether a global procedure like memory test A which explicitly reflects these critical influences by (a) asking for additional information and (b) asking for the sources of remembered pieces of information can help avoid these potential risks. Of course, it is certainly not easy to transfer this procedure to applied issues, but research directed explicitly at the relevant conversion processes might identify crucial variables that could improve real testimonies.

Final remarks

If there is something like a quintessence of this work, then it is a perspective change emerging from it. This change is - in the terminology of part I - from impairment effects to transformational effects or - in Tulving's terms - from ecphory to conversion processes. For a long time, it has been a guiding assumption that the impact of additional, discrepant information has to do with changes in memory - of whatever kind, overwriting, integration, inacessibility, failures of source discrimination, and so on. As a consequence, situational factors were assigned only a supplementary role, that is, they were built into the assumed memory mechanisms as mediating influences. For instance, the impact of warnings has been interpreted as overcoming a blocking mechanism (Christiaansen & Ochalek, 1983) or preventing the integration of misinformation (if presented prior to its encoding; Greene, Flynn, & Loftus, 1982). Furthermore, as already noted in the introduction, the suggestion of misleading details via presuppositions was seen as merely setting the stage for its encoding without calling too much attention to it, thus allowing for its integration (Loftus, 1975). Thus, memory has been the frame for theorizing and conducting research. Even McCloskey and Zaragoza's (1985) demonstration of responses biases (i.e., transformational processes) with the standard test procedure and subsequent introduction of the modified test procedure aimed at a proper assessment of memory, not at an investigation of these transformational processes per se.

In short, the default assumption has been that something special is happening in memory as a result of introducing misinformation, and the precise way it happens depends on situational factors. We could term this a *memory perspective*. Of course, there is nothing wrong per se with this perspective. My argument is merely - and this is a commonplace observation in human problem solving - that it is sometimes profitable to change one's perspective of a problem in order to arrive at a solution. Why not approach the misinformation effect "from without" - that is, begin with the test situation and then ask what it makes of various memory constellations? Thus, a *situational perspective* could assume by default that nothing special happens in memory and place the explanatory burden on aspects of the situation in which the remembering takes place. Two such situational aspects have been adressed in this work: the influence of conversational maximes in part I and the utilization of local vs. global test procedures in part II.

A situational perspective is particularly attractive from a practical standpoint: Situational variables under the control of the investigator could be used to improve the performance of eyewitnesses. A particularly good illustration comes from recent research on lineup identification which has uncovered a bulk of situational variables that bear heavily on rates of correct and false identifications. For example, merely telling eyewitnesses that the actual culprit might or might not be present in the lineup diminishes the rate of false identifications in lineups where the culprit is absent, with no change in the rate of correct identifications when the culprit is

present (Malpass & Devine, 1981). Also, it has repeatedly been found that sequential presentation of lineups leads to fewer false identifications (with no loss in correct identifications) than a simultaneous presentation (see Wells, 1993). Thus, it seems to be possible to create circumstances that make the best out of people's memories. In the case of lineup identifications, this was possible because researchers had a theory about the processes that guide eyewitnesses' identification decisions: it was held that eyewitnesses make relative judgments, that is, they tend to pick out the person that most closely matches their memory of the perpetrator, compared to other members of the lineup (Wells, 1984). On the basis of this theoretical assumption, it seemed reasonable that sequential presentation of lineups, for instance, should work better because it precludes relative judgments. In a similar fashion, it should be possible to derive safeguards against the impact of misleading post-event information. The theoretical framework as well as the two manipulations addressed in this work, enlightenment and global testing, may serve as starting points for the development of such procedures. However, these measures can most probably not be directly applied to real testimonies since their effectiveness depends to a considerable degree on the experimenter's knowledge that the subjects received discrepant information from particular sources of information. In most real-life situations, we have no certain knowledge that and where misinformation was available. Hence, a necessary step towards the development of "misinformation safeguards" would be to assure the effectiveness of such procedures under these conditions.

Adopting a situational perspective seems also reasonable because some of the memory effects that have indeed been discovered in eyewitness research are effects that depend heavily on changes in the retrieval situation; I mean context effects. However, the evidence concerning such effects is far from being unequivocal. For example, Bekerian and Bowers (1983) have found that presenting the memory test questions in the same sequential order than the original slide sequence drastically improved performance for the original details. On the other hand, McCloskey and Zaragoza (1985) failed to find such facilitatory effects. With respect to mode of testing, Pezdek and Greene (1993) found that visual testing markedly improved performance compared to verbal testing. This is a rather untypical finding; Loftus, Miller, and Burns (1978), for instance, also used a visual test and obtained much larger misinformation effects than the former authors, and generally, the amount of the misinformation effect across studies is rather insensitive to mode of testing. Thus, whether and to what extent context reinstatement works remains an open question. From a situational perspective, the first thing to do would be to look for - perhaps minor - differences in the retrieval situations that may eventually turn out to be responsible for the conflicting results. This could end up in finding a classification of retrieval cues and, perhaps more important, in identifying a pattern of relationships between these cues: Some may affect memory performance independently from each other, whereas for others a hierarchical relationship may hold; that is, a given cue's effectiveness rests upon the presence of another cue.

--- 87 ---

In the end, this research logic might even lead to a new conception of memory impairment. Following Loftus (1991), the notion of memory impairment can be subdivided into trace impairment and retrieval impairment. Retrieval impairment means minor acessibility due to context effects, as discussed above. From a situational perspective, however, one might ask whether there is any room left at all for *trace* impairment. The reason for doubting this is theoretical. Trace impairment is, by definition, tied to the concept of the memory trace; we might say, for instance, that impairment means a decrement in the strength of the trace. Since the trace is assumed to be laid down somewhere in the person, it follows that its attributes, like strength, should be independent of the situation wherein it is used. Now, however, we run into the principal difficulty that any conclusions about memory traces must be inferred from people's behavior in concrete situations, for the simple reason that memory traces are not (at least, not yet) directly observable. This difficulty led some researchers, like Watkins (1990), to abandon the concept of memory traces altogether (though he does not deny the existence of memory traces but argues that they are beyond the reach of the psychologist's techniques). Instead, he advocates the formulation of empirical laws - like the cue overload principle (Watkins, 1979) - that tie aspects of retrieval situations to memory performance and/or encoding conditions.

Thus, it may turn out that trace impairment is no more or no less open to investigation than retrieval impairment. What we can say about the former is the same that we can say about the latter: what aspects of situations are responsible for its occurrence. Maybe the difference is merely that in one case these aspects are temporally more stable, thus leaving the impression that some enduring damage has been done, or that more hints are necessary for successful performance. In any case, the new conception of memory impairment would be in terms of the restrictions that have to be put on the retrieval situation in order that experimental performance equals control performance. Needless to say that such a systematic analysis of situational determinants of memory performance - if it could be done - would be much more valuable than ideological debates on presumed degrees of memory impairment.

[1]Actually, this construction seems a little odd; one might say that memory states 2 as well as 4 are not possible to exist at all, respectively, that they "turn into" memory states 1 for practical purposes. However, this would not change the results of the performance calculations. Therefore, and to avoid conceptual complications, I want to retain these memory states, knowing that they are "virtual" ones. Furthermore, they aquire reasonable existence in designs with more than two sources of information, like in the experiments reported here.

[2]One could also make more sophisticated assumptions, for example, that subjects adjust their decisions to the (objective or subjectively remembered) frequencies of identical and different response terms. However, this is not necessary for purposes of the present simulations.

[3]Optimal uncertainty concerning the correspondence of first and second list terms would result from equal numbers of consistent and different responses; this, however, would require subjects to learn a large amount of consistent response terms which are of no interest of their own; these resources could better be used to raise power for the tests in other conditions. Therefore I contented myself with a smaller amount of consistent reponse terms, hoping that the subjects would not be too sensitive to the actual frequencies and adjust their guesses accordingly, admittedly a questionable assumption.

[4]Note that though memory test A covered source memory as well, the data reported in this table do not take source memory into account; what counts is only correct recognition of response terms, regardless whether they are correctly attributed to lists or not.

[5]This already indicates that the independence assumption mentioned in the introduction of the model does not hold. However, this does not touch the validity of the following arguments. - Apart from that, a word on the statistical analysis is in order. While the emphasis is on the simulation of the test B results (see below), one might nevertheless argue that some kind of analysis of variance is appropriate since it was predicted that there are RI and PI effects in test B but not necessarily in test A (or, at least, that there are different RI and PI effects in these tests, if the independence assumption does not hold). In either case, some kind of interaction should be expected. This was partially confirmed in a within-subjects ANOVA with condition (experimental vs. control) and test (A vs. B) as fixed effects: There was a significant interaction for RI, $F(1, 27) = 24.30$, $p < .001$, but not for PI, $F(1, 27) = 1.62$, $p = .21$. The main effects per se are uninteresting; therefore, the more illustrative results of the planned comparisons (amount of RI or PI for each test) are indicated in table 2 by asterisks.

[6]This explanation was suggested independently by Wolfgang Hell and one of my subjects. Note that I observed this reversed RI effect not only in experiment 1 but also in two small pilot studies with longer retention intervals and a recall procedure. In these studies, experimental performance was sometimes twice as high as control performance (though on a very low overall level, i.e., around 20% - 30% correct recall).

[7]The same ANOVA as performed on the experiment 1 results yielded a significant interaction effect of test and condition for RI, $F(1, 27) = 5.35$, $p < .05$, but not for PI, $F(1, 27) = .54$, p =

References

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

Adler, J. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behavior, 14,* 165-181.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203-231.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58,* 97-105. Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, Eng.: Cambridge Univ. Press.

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

Bekerian, D. A., & Dennett, J. L. (1990). Spoken and written recall of visual narratives. *Applied Cognitive Psychology, 4,* 175-187.

Belli, R. F. (1988). Color blend retrievals: Compromise memories or deliberate compromise responses? *Memory & Cognition, 16,* 314-326.

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

Belli, R. F., Windschitl, P. D., McCarthy, T. T., & Winfrey, S. E. (1992). Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 356-367.

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 and Cognition, 22,* 40-54.

Blank, H. (1994). Täuschung, Aufklärung und der Falschinformationseffekt in Zeugenaussagen-Experimenten. In S. L. Sporer & D. Meurer (Eds.), *Die Beeinflußbarkeit von Zeugenaussagen* (pp. 103-127). Marburg: Elwert.

Blank, H., Wienholt, H., & Sporer, S. L. (1995, April). *Gibt es Langzeit-Verfälschungen von Zeugenerinnerungen als Folge früherer Erinnerungstests?* Paper presented at the 37. Tagung experimentell arbeitender Psychologen in Bochum, Germany.

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

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

Bowman, L. L., & Zaragoza, M. S. (1989). Similarity of encoding context does not influence resistance to memory impairment following misinformation. *American Journal of Psychology, 1989,* 249-264.

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General, 116,* 38-49.

Chandler, C. C. (1989). Specific retroactive interference in modified recognition tests: Evidence for an unknown cause of interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 256-265.

Chandler, C. C. (1991). How memory for an event is influenced by related events: Interference in modified recognition tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 115-125.

--- 89 ---

Chandler, C. C. (1993). Accessing related events increases retroactive interference in a matching recognition test. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 967-974.

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.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, W. G., & Loftus, E. F. (1979). Incorporating new information into memory. *American Journal of Psychology, 92,* 413-425.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

Davis, J., & Schiffman, H. R. (1985). The influence of the wording of interrogatives on the accuracy of eyewitness recollections. *Bulletin of the Psychonomic Society, 23,* 394-396.

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

Dodson, C., & Reisberg, D. (1991). Indirect testing of eyewitness memory: The (non)effect of misinformation. *Bulletin of the Psychonomic Society, 29,* 333-336.

Edwards, D., & Middleton, D. (1986a). Joint remembering: Constructing an account of shared experience through conversational discourse. *Discourse Processes, 9,* 423-459.

Edwards, D., & Middleton, D. (1986b). Text for memory: Joint recall with a scribe. *Human Learning, 5,* 125-138.

Franzen, S., & Sporer, S. L. (1994). Personenverwechslungen durch irreführende Rekonstruktionsbilder: Zum Einfluß nachträglicher Informationen und der Wiederherstellung des Wahrnehmungskontextes. In S. L. Sporer & D. Meurer (Eds.), *Die Beeinflußbarkeit von Zeugenaussagen* (pp. 207-236). Marburg: Elwert.

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

Gibling, F., & Davies, G. (1988). Reinstatement of context following exposure to post-event information. *British Journal of Psychology, 79,* 129-141.

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

Greeno, J. G., James, C. T., & DaPolito, F. J. (1971). A cognitive interpretation of negative transfer and forgetting of paired associates. *Journal of Verbal Learning and Verbal Behavior, 10,* 331-345.

Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107,* 311-327.

Hays, W. L. (1973). *Statistics for the social sciences.* London: Holt, Rinehart and Winston.

Heine, R. (1914). Über Wiedererkennen und rückwirkende Hemmung. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 68,* 161-236.

Hertwig, R., & Gigerenzer, G. (1994). *The chain of reasoning in the conjunction task.* Unpublished manuscript.

Higgins, E. T., & Stangor, C. (1988). A "change-of-standard" perspective on the relations among context, judgment, and memory. *Journal of Personality and Social Psychology, 54,* 181-192.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities.* Reading, MA: Addison-Wesley.

Holtgraves, T., Srull, T. K., & Socall, D. (1989). Conversation memory: The effects of speaker status on memory for the assertiveness of conversation remarks. *Journal of Personality and Social Psychology, 56,* 149-160.

Howe, M. L. (1991). Misleading children's story recall: Forgetting and reminiscence of the facts. *Developmental Psychology, 27,* 746-762.

Kallio, K. D., & Cutler, B. L. (1987). Does lexical marking affect eyewitness recall? *Journal of Psychology, 121,* 249-258.

Köhnken, G., & Brockmann, C. (1987). Unspecific postevent information, attribution of responsibility, and eyewitness performance. *Applied Cognitive Psychology, 1,* 197-207.

Kroll, N. E. A., Ogawa, K. H., & Nieters, J. E. (1988). Eyewitness memory and the importance of sequential information. *Bulletin of the Psychonomic Society, 26,* 395-398.

Kroll, N. E. A., & Timourian, D. A. (1986). Misleading questions and the retrieval of the irretrievable. *Bulletin of the Psychonomic Society, 24,* 165-168.

Lassiter, G. D., Stone, J. I., & Rogers, S. L. (1988). Memorial consequences of variation in behavior perception. *Journal of Experimental Social Psychology, 24,* 222-239.

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

Lipscomb, T. J., McAllister, H. A., & Bregman, N. J. (1985). Bias in eyewitness accounts: The effects of question format, delay interval, and stimulus presentation. *Journal of Psychology, 119,* 207-212.

Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology, 62,* 90-95.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology, 7,* 560-572.

Loftus, E. F. (1977). Shifting human color memory. *Memory & Cognition, 5,* 696-699.

Loftus, E. F. (1979a). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. F. (1979b). Reactions to blatantly contradictory information. *Memory & Cognition, 7,* 368-374.

Loftus, E. F. (1981). Mentalmorphosis: Alterations in memory produced by the mental bonding of new information to old. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 417-434). Hillsdale, NJ: Erlbaum.

Loftus, E. F. (1991). Made in memory: Distortions in recollection after misleading information. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, 25* (pp. 187-215). New York: Academic Press.

Loftus, E. F., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition, 17,* 607-616.

Loftus, E. F., & Greene, E. (1980). Warning: Even memory for faces may be contagious. *Law and Human Behavior, 4,* 323-334.

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General, 118,* 100-104.

Loftus, E. F., Levidow, B., & Duensing, S. (1992). Who remembers best? Individual differences in memory for events that occurred in a science museum. *Applied Cognitive Psychology, 6,* 93-107.

Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist, 35,* 409-420.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19-31.

Loftus, E. F., & Palmer, J. P. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585-589.

Loftus, E. F., Schooler, J. W., & Wagenaar, W. A. (1985). The fate of memory: Comment on McCloskey and Zaragoza. *Journal of Experimental Psychology: General, 114,* 375-380.

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society, 5,* 86-88.

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 343-350.

Martin, E. (1971). Verbal learning theory and independent retrieval phenomena. *Psychological Review, 78,* 314-332.

Mayring, P. (1988). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Deutscher Studienverlag.

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

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39,* 352-370.

McSpadden, M. D., Schooler, J. W., & Loftus, E. F. (1988). Here today, gone tomorrow: The appearance and disappearance of context effects. In G. M. Davies and D. M. Thomson (Eds.), *Memory in Context: Context in Memory* (pp. 215-229). Chichester: John Wiley.

Mélen, M. (1989). L'effet de la présentation d'une information nouvelle erronée: une synthèse des travaux récents. *Les Cahiers Internationaux de Psychologie Sociale, 1,* 9-29.

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology, 53,* 173-203.

Mensink, G., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434-455.

Metcalfe, J. (1990). Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony. *Journal of Experimental Psychology: General, 119,* 145-160.

Moston, S. (1990). How children interpret and respond to questions: Situational sources of suggestibility in eyewitness interviews. *Social Behaviour, 5,* 155-167.

Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, Ergänzungsband 1.* Leipzig: Barth.

Niemi, P. (1979). A note on "blocking" as a retrieval interference phenomenon. *American Journal of Psychology, 92,* 547-550.

Norman, D. A. (1970). Comments on the information structure of memory. In A. F. Sanders (Ed.), *Attention and Performance III* (pp. 293-303). Amsterdam: North Holland.

Ober, B. A., & Stillman, R. C. (1988). Memory in chronic alcoholics: Effects of inconsistent versus consistent information. *Addictive Behaviors, 13,* 11-15.

Pezdek, K., & Greene, J. (1993). Testing eyewitness memory: Developing a measure that is more resistant to suggestibility. *Law and Human Behavior, 17,* 361-369.

Pohl, R. F., Schumacher, S., & Friedrich, M. (1993). The eyewitness-misinformation effect: Distorted recollections based on contradictory information. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 33-52). Elsevier Science Publishers B. V.

Postman, L. (1961). The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior* (pp. 152-179). New York: McGraw-Hill.

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

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior, 7,* 672-694.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory and Cognition, 1,* 19-40.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93-134.

Read, J. D., & Bruce, D. (1984). On the external validity of questioning effects in eyewitness testimony. *International Review of Applied Psychology, 33,* 33-49.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96,* 341-357.

Rudy, L., & Goodman, G. S. (1991). Effects of participation on children's reports: Implications for children's testimony. *Developmental Psychology, 27,* 527-538.

Ryan, R. H., & Geiselman, R. E. (1991). Effects of biased information on the relationship between eyewitness confidence and accuracy. *Bulletin of the Psychonomic Society, 29,* 7-9.

Saywitz, K. J., Goodman, G. S., Nicholas, E., & Moan, S. F. (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology, 59,* 682-691.

Schmidt, R. (1987). "Über Wiedererkennen und rückwirkende Hemmung": Neuauswertung einer Arbeit von Rosa Heine, 1914. *Zeitschrift für Psychologie, 195,* 223-230.

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 171-181.

Schooler, J. W., & Tanaka, J. W. (1991). Composites, Compromises, and CHARM: What is the evidence for blend memory representations? *Journal of Experimental Psychology: General, 120,* 96-100.

Schumacher-Bittner, S. (1994). Modalitätseinflüsse auf Falschinformationseffekt und Quellenüberwachung. In S. L. Sporer und D. Meurer (Eds.), *Die Beeinflußbarkeit von Zeugenaussagen* (pp. 128-153). Marburg: Elwert.

Schuurmans, E., & Vandierendonck, A. (1985). Recall as communication: Effects of frame anticipation. *Psychological Research, 47,* 119-124.

Shaughnessy, J. J., & Mand, J. L. (1982). How permanent are memories for real life events? *American Journal of Psychology, 95*, 51-65.

Sheehan, P. W. (1988). Memory distortion in hypnosis. *International Journal of Clinical and Experimental Hypnosis, 36,* 296-311.

Sheehan, P. W., & Grigg, L. (1985). Hypnosis, memory, and the acceptance of an implausible cognitive set. *British Journal of Clinical and Experimental Hypnosis, 3,* 5-12.

Sheehan, P. W., Grigg, L., & McCann, T. (1984). Memory distortion following exposure to false information in hypnosis. *Journal of Abnormal Psychology, 93,* 259-265.

Sheehan, P. W., & Statham, D. (1989). Hypnosis, the timing of its introduction, and acceptance of misleading information. *Journal of Abnormal Psychology, 98,* 170-176.

Sheehan, P. W., & Tilden, J. (1983). Effects of suggestibility and hypnosis on accurate and distorted retrieval from memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 283-293.

Sheehan, P. W., & Tilden, J. (1984). Real and simulated occurrences of memory distortion in hypnosis. *Journal of Abnormal Psychology, 93,* 47-57.

Sheehan, P. W., & Tilden, J. (1986). The consistency of occurrences of memory distortion following hypnotic induction. *International Journal of Clinical and Experimental Hypnosis, 34,* 122-137.

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.

--- 93 ---

Sporer, S. L., & aus dem Kahmen, A. (1994). Irreführung von Zeugen durch Falschinformationen: Empirische Befunde zur Beeinflußbarkeit von Aussagen. In S. L. Sporer und D. Meurer (Eds.), *Die Beeinflußbarkeit von Zeugenaussagen* (pp. 58-81). Marburg: Elwert.

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

Trouvé, R. J., & Libkuman, T. M. (1992). Eyewitness performance of personality types as a function of induced arousal. *American Journal of Psychology, 105,* 417-433.

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

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology, 21,* 437-484.

Tversky, B., & Tuchin, M. (1989). A reconciliation of the evidence on eyewitness testimony: Comments on McCloskey and Zaragoza. *Journal of Experimental Psychology: General, 118*, 86-91.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64,* 49-60.

Vandierendonck, A. (1986). Het schema-begrip in de hedendaagse geheugentheorie. *Psychologia Belgica, 26,* 161-183.

Wagenaar, W. A., & Boer, J. P. A. (1987). Misleading postevent information: Testing parameterized models of integration in memory. *Acta Psychologica, 66*, 291-306.

Wagenaar, W. A., & Loftus, E. F. (1990). Ten cases of eyewitness identification: Logical and procedural problems. *Journal of Criminal Justice, 18,* 291-319.

Ward, R. A., & Loftus, E. L. (1985). Eyewitness performance in different psychological types. Journal of General Psychology, 112, 191-200.

Warren, A., Hulse-Trotter, K., & Tubbs, E. C. (1991). Inducing resistance to suggestibility in children. *Law and Human Behavior, 15,* 273-285.

Watkins, M. J. (1979). Engrams as cuegrams and forgetting as cue overload: A cueing approach to the structure of memory. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 347-372). New York: Academic Press.

Watkins, M. J. (1990). Mediationism and the obfuscation of memory. *American Psychologist, 45,* 328-335.

Weinberg, H. I., Wadsworth, J., & Baron, R. S. (1983). Demand and the impact of leading questions on eyewitness testimony. *Memory & Cognition, 11,* 101-104.

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89-103.

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48,* 553-571.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71,* 291-301.

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

Zaragoza, M., & McCloskey, M. (1989). Misleading postevent information and the memory impairment hypothesis: Comment on Belli and reply to Tversky and Tuchin. *Journal of Experimental Psychology: General, 118*, 92-99.