Can private reports enhance children’s event recall, lower their suggestibility and foster their metacognitive monitoring compared to face-to-face interviews?

Claudia M. Roebers a,*, Pauline Howie b, Eva Beuscher a

a Department of Psychology IV, University of Würzburg, Roentgenring 10, 97070 Würzburg, Germany
b Department of Psychology A 17, University of Sydney, NSW 2006, Australia

Available online 8 December 2004

Abstract

In this paper, two studies are presented in which the social demands of an event recall interview situation were manipulated using two different methodologies and by using different dependent measures as indices for children’s event memory, suggestibility and metacognitive monitoring processes. Participants aged 6–10 years were shown a brief video and then questioned about it 2–3 weeks later. Answers to the recall questions were either given privately by whispering into a teddy bear’s ear or publicly in a normal face-to-face interview. Confidence judgments as indicators for metacognitive monitoring were entered into a computer, either with the interviewer having direct knowledge about them, or with the children entering the confidence judgments in privacy. In line with previous research on this topic, although private reports did not produce poorer performance, neither did they improve performance. In both studies there were no effects of the social manipulations on either recall or metacognitive monitoring and also no age-dependent benefits. The results are discussed in terms of the chosen methodologies and in the light of previous studies.

Keywords: Children; Interview; Event recall; Development
1. Introduction

There is now a large body of literature on children’s event recall and suggestibility consistently showing that, in addition to chronological age, a variety of cognitive factors contribute to the developmental progression in the amount of information children can correctly report about a past event and in their resistance to false suggestions. For example, efficiency of encoding (Marche, 1999), prior knowledge (Ornstein, Shapiro, Clubb, Follmer, & Baker-Ward, 1997), metacognitive monitoring processes (Roebers, 2002), source attribution (Roberts & Blades, 2000; Zaragoza & Lane, 1994), and the underlying memory trace (Pedzek & Roe, 1995) have been shown to play an important role in explaining performance differences between children within homogeneous age groups.

In addition to the understanding that cognitive factors affect event memory and suggestibility in children, there is also growing consensus about the importance of social factors. Research has addressed the question of whether, to what extent, and by what mechanisms, children’s memory reports and suggestibility are influenced by the social demands of the interview situation in which the memory report is sought. There is growing evidence of the influence of the person of the interviewer (Goodman, Sharma, Thomas, & Considine, 1995; Jackson & Crockenberg, 1998), the status and/or familiarity of the interviewer (Bjorklund et al., 2000; Cassel & Bjorklund, 1995), warning instructions about tricky questions (Warren et al., 1999), repetition of questions (Howie, Sheehan, Mojarrad, & Wrzesinska, in press; Poole & White, 1995), and question wording and format (Gee, Gregory, & Pipe, 1999; Roebers & Howie, 2003).

Such studies have shown that children are sensitive to the social demands of the interview situation, and may perceive the social situation differently from adults. Moreover, there are developmental differences, with older children apparently having much more knowledge about social interactions and communication rules than younger children (Saywitz, 2002; Schwarz, Roebers, & Schneider, 2004; Siegal, 1991; Waterman, Blades, & Spencer, 2000). In fact, it appears that young children feel an obligation to answer nearly every question. Typically, younger children show low frequencies of “I don’t know” responses and a high number of incorrect answers to specific but unbiased questions. Additionally, when questions are misleading, young children have a strong tendency to go along with the interviewer’s (false) suggestions instead of either rejecting the suggestion or skipping a question and its answer by using the “I don’t know” option – a strategy used by adults (Roebers and Schneider, 2000). Thus, certain question formats can generate social demands in an interview situation, which can undermine children’s otherwise competent performance (for example, children’s superior recognition performance compared to their performance on specific, biased or misleading questions).

The evidence that perceived social demands affect children’s performance is consistent with the documented benefits of interviewer-provided social support on children’s resistance to suggestions. Although the existing literature is small, the evidence points to a positive effect of a warm interview atmosphere (for example, use of
eye contact, smiling, and a warm voice) that builds rapport and fosters the child’s feeling of well-being and self-efficacy (Carter, Bottoms, & Levine, 1996; Davis & Bottoms, 2002; Goodman, Bottoms, Rudy, Davis, & Schwartz-Kenney, 1991; Roberts, Lamb, & Sternberg, 2004). Although such interviewer-provided social support appears to decrease children’s resistance to suggestions, it does not necessarily enhance their free reports (Davis & Bottoms, 2002). These findings, taken together with the findings on “I don’t know” responses, provide persuasive evidence that a reduction of social demands has positive effects on children’s answers to specific questions, especially when misleading.

Another method of decreasing the social demands of an interview situation may be to use a computer as interviewer. Interviewing with the help of a computer has proved useful in eliciting personal, sensitive, or embarrassing information from adolescents and adults in clinical settings such as drug-use or sexual abuse cases (Bagley & Genius, 1991; Romer et al., 1997). Moreover, when questioned about their preference for personal versus computerized interviews, the majority of adults and adolescents have consistently acknowledged a preference for computerized interviews, especially when the interview content is personal. Together, these findings have been interpreted as indicating that computers create a private and confidential atmosphere, are perceived to be less demanding and allow the respondent greater control than with human interviewers. These perceptions are argued to generate a more truthful atmosphere, conducive to a reduced likelihood of socially desirable answers (Paperny, Aono, Lehmann, Hammar, & Risser, 1990). There is, however, limited information about children’s responses to interviews by computers or other non-human methods, both when the interview content is sensitive and personal, and when it is more neutral. The focus of the studies that will be reported here is on interviews about neutral content, for two reasons. Firstly, it is important to investigate the possibility that children’s responses to a non-human “interviewer” might actually reduce the information they provide. Secondly, in real cases of interrogation of children, it is often not known whether and when the areas covered are going to be very personal or embarrassing. Therefore, effects of non-human interviews with children need to be investigated.

It is also possible that non-human/computerized interviews may be of use in counteracting the high frequency of incorrect answers to specific, unbiased questions in younger children and their heightened suggestibility in typical recall interview contexts (see Bruck & Ceci, 1999 for a review). Such interviews may decrease any perceived obligation to respond to questions even when ignorant of the answer, simply in order to please the interviewer (e.g., Poole & Lamb, 1998; Waterman et al., 2000). Similarly, in the case of leading questions, such interviews may reduce children’s temptation to yield to misleading suggestions because they are reluctant to contradict the interviewer’s suggestion, or because they believe that the interviewer must be right. Empirical research on this question is still in its very early stages, however. Steward et al. (1996) were the first to investigate the possible benefits of a computerized interview in young children. With the help of anatomically detailed dolls presented either on a computer or on paper, 3- to 6-year-olds were questioned face-to-face about a pediatric examination which had taken place some days earlier.
There were no differences between the computer and paper condition in the amount of accuracy of children’s reports, regardless of age. However, it was only the presentation of recall prompts that was computerized and thus, the findings are not very surprising.

To date, only two studies have investigated the effects of an entirely computerized interview on children’s event recall. Donohue, Powell, and Wilson (1999) developed an interactive computer program to assess children’s event recall. Participants were questioned about various special activities they had carried out in their classroom with their teacher 3–4 days earlier. There were three interview conditions in the study: (a) a standard interview condition with an adult interviewer, (b) a computerized interview in which a “human” face portrayed on the computer monitor “spoke” to the child and the child responded into a microphone attached to the computer, and (c) a computerized interview as in (b) but with an adult interviewer present throughout the interview to assist the child. Dependent measures were the number of correct recognition responses to yes/no questions about whether particular activities had occurred, the number of items of information recalled in response to very broad, open-ended prompts about each activity recognized, and finally disclosure of the “secret” that had been introduced during the target event. There were no differences between any of the interview conditions with respect to the number of correctly recalled features, intrusion errors or confusions, regardless of age. There were also no significant differences between conditions in the number of children who kept or disclosed the secret. However, children seemed to engage more closely with the computerized task than children in the control conditions because they showed greater willingness to go back and revise responses. In spite of this, the revision did not benefit either amount or accuracy of recall, since most of the information given in revisions had already been mentioned in the initial responses.

In a follow-up study using similar recall measures and the same conditions, Powell, Wilson, and Thomson (2002) applied a more stringent coding for secret disclosure, and also measured response latencies with the aim of providing a more sensitive measure of the effect of the computer interview. Additionally, recall interviews were conducted on two occasions (after 3–4 days and again after 14 days). Overall, the study confirmed the previous findings in that no differences between the conditions were found, either in correct recognition of event features, in accurate and inaccurate recall of details, or in correct rejection of activities that had not occurred. Contrary to expectation, in the human interview condition there were more secret disclosures than in the computer condition, but children took longer to respond to recognition questions in the human condition. As in the earlier study, children in the computer condition opted to revise more questions, but again these revisions did not significantly increase the correct information recalled. Also, the increase in revisions was no longer present in the second computer interview in the younger age group and was therefore interpreted as a novelty effect associated with the computer task. Compared with the human interview, the computer interview showed less consistency across the two recall interviews, but this was not reflected in any changes in accuracy over time.
Although the results of these studies are quite striking in that computer interviews in the absence of an adult do not undermine the amount or accuracy of children’s reports, they show little or no evidence that computers improve these aspects of recall. However, it is premature to conclude at this stage that “private” reporting which is indirect and non-confrontational is of no value. First, it is important that the children appeared to enjoy the task. Second, the delay before the first interview was very short (3–4 days) so that memory decay may have been minimal, and it appears that the recognition questions were not difficult as on average at least 70% were answered correctly. With longer delays and with more difficult material the manipulation may yield different results. Third, the recall measures used in these studies are known to put relatively little social pressure on children. Responses to yes/no recognition questions are generally considered to be a relatively good indicator of “pure” underlying memory (Schneider & Bjorklund, 1998), and although questions to which the correct answer is “no” may be subject to some perceived pressure to acquiesce, the level of social pressure is low. Furthermore, the open-ended prompts used in these studies resemble free recall tasks and free recall has been shown to elicit minimal – although accurate – recall in young children (e.g., Bruck, Melnyk, & Ceci, 2000; Elischberger & Roebers, 2001; Poole & Lamb, 1998; Roebers & Schneider, 2001a) and to be unaffected by interviewer-provided social support (Davis & Bottoms, 2002). Specific and leading questions, in contrast, have repeatedly been shown to put high social demands on the children and therefore to be more prone to inaccurate responding and to be significantly affected by the presence or absence of social support (Carter et al., 1996; Goodman et al., 1991).

Thus, it is possible that benefits of non-face-to-face interviews will emerge more clearly in children’s answers to specific questions and false suggestions, since these are likely to be more sensitive to any effects of decreased social demands in such interviews. It is also possible, however, that any benefits of private reporting will be counteracted by the impersonal nature of speaking to a computer. Thus, it may be necessary to devise other means of private reporting which are less impersonal. The two studies presented here explore various aspects of private reporting, in an attempt to begin to identify what, if any, benefits accrue, under what conditions and reflecting what underlying mechanisms.

In our first study, we compared children’s event recall accuracy and suggestibility in a standard face-to-face interview (public recall) and a private interview in which children whispered the answer to each question into a teddy bear’s ear (private recall). We also included question formats more likely to involve social pressure to “please” the interviewer, namely strongly misleading questions to which the correct answer is “no” and which implicitly convey the interviewer’s expected answer, as well as specific, though unbiased questions. We hypothesized that a teddy bear would more easily be perceived to be a friend to whom honest answers can be given even if they are not socially desirable. Although the earlier studies found no age differences in the effects of computerized interviews, we included 3 age groups (6, 7 and 8 years) since strong improvements with respect to children’s ability to resist false suggestions and to adequately monitor their answering behavior concerning
unbiased questions are typically observed in this phase of development (Roebers & Fernandez, 2002; Roebers, Moga, & Schneider, 2001). In contrast to the earlier studies, participants in the present study watched a brief video and were questioned about it 2–3 weeks later.

We expected that in the private, whispering condition, there would be reduced social pressure in the form of perceived expectations to please the interviewer, and that children would therefore be more resistant to misleading questions and more accurate when answering specific questions. However, we expected the effect to be smaller for unbiased questions because for this question format the social demands are less strong and concern the ability to withhold uncertain answers rather than the ability to resist false suggestions.

2. Study 1

2.1. Method

2.1.1. Overview

A 3 (age) × 2 (question format) × 2 (response format) factorial design was used in the study. Children aged 6, 7 and 8 years watched a brief video and were individually questioned about it. In the course of the interview, two different question formats (unbiased open-ended and misleading questions) were used. Furthermore, half of the children were asked to whisper the answer in a teddy bear’s ear (private response format), whereas the children in the public response format gave their answers directly to the interviewer.

2.1.2. Sample

A total of 122 participants (71 girls, 51 boys) from three age groups were included in the study: 33 6-year-olds (mean age = 6.4; SD = 4 months); 45 7-year-olds (mean age = 7.5; SD = 4 months); and 44 8-year-olds (mean age = 8.4; SD = 5 months). Children were recruited from three different kindergartens and two different elementary schools in a suburban area of Bavaria (Germany) and came from lower to upper middle-class families. Sixteen 6-year-olds were randomly assigned to the public response format (8 girls) and 17 to the private condition (10 girls). Of the 7-year-olds, 23 were assigned to the public response format (12 girls) and 22 to the private condition (14 girls). Finally, 22 of the 8-year-olds were assigned to the public response format (13 girls) and 22 to the private answer condition (14 girls).

2.1.3. Procedure

Participants were shown a short video (7 min) in small groups. Prior to watching the video, participants were not informed of the purpose of the study, to prevent them from deliberately memorizing the event. Instead, they were simply told that the experimenter wanted their opinion about the video and they were asked to pay full attention. After the video, participants were asked whether they enjoyed
the video. About three weeks later ($M = 23$ days, $SD = 4$ days), the children were individually interviewed in a quiet room in their school or kindergarten by an experimenter who was not known to them, and who had not presented the video. The interview lasted about 15 min. Answers to all questions were either recorded on standard protocol sheets for later coding (public response format) or recorded on a mini-disc player with a microphone concealed in the teddy bear’s ear (private response format). In both public and private response formats, the big teddy bear was positioned on the table before the child’s arrival, and used to build rapport with the child. After establishing rapport, the interviewer asked the child to report as much as possible about the video, and then asked 22 standard questions. Children were instructed to answer with “I don’t know” whenever they were not sure about the right answer. They were, however, not especially warned about some questions being misleading. The children in the private response format were asked to whisper the answer into the teddy bear’s ear so that the interviewer would not hear it. (The interviewer explained to the child that she had not seen the film yet and did not want to spoil it by finding out what happened before seeing it). A wooden wall between the interviewer and the teddy bear made it clear to the children that the interviewer could not hear their answer (which was in fact the case). Before the interview began, the child was trained to whisper to the bear using modeling if necessary (i.e., if the child’s whispering could not be detected by the recorder). During the questioning, none of the children received feedback about whether their answers were correct. In order to check the experimental manipulation, after the interview was completed, the children in the private response format were asked if they believed that the bear would keep their answers confidential. The children were also asked not to tell other children about the video and the interview. They were praised for their good performance, given a small gift and returned to their classroom.

2.2. Materials

2.2.1. Stimulus

The video is called “Treasure Hunt”, and depicts a group of children who visit a farmer’s family during their vacation. The children learn that a castle once stood on the farm but was destroyed by a fire. They search for the ruins of the castle, find the former entrance to the dungeons, climb down into underground tunnels and eventually discover treasure.

2.2.2. Questions

The event recall task involved 22 questions: 6 filler questions and 16 target questions. For each participant, 8 target questions were asked in an unbiased, open-ended format and 8 in a strongly misleading format. Questions in the misleading format strongly suggested an incorrect answer (e.g., “The boys came to the farmhouse by car, right?”) and participants were asked to respond with either “yes” or “no”. In each case, “no” was the correct answer. The questions in open-ended format (unbiased) focused on the same details as the misleading questions but did not
suggest a specific answer (e.g., “How did they boys get to the farmhouse?”). Both the misleading and unbiased questions have been used in previous studies (Roebers & Fernandez, 2002; Roebers & Howie, 2003; Roebers & Schneider, 2002). Two sets of 22 questions were created, each with 8 unbiased questions, 8 misleading questions and 6 filler questions which were asked in a positive leading question format, that is, they suggested the correct answer. The 6 filler questions were identical in both sets, and were interspersed among the target questions in order to reduce the uniformity of question format and to maintain the interviewer’s credibility. The 8 questions asked in unbiased form in Set 1 were asked in misleading form in Set 2, and vice versa. Sixty-four participants were questioned with Set 1, and 58 with Set 2.

2.3. Results

All main effects and interactions are reported with $p < 0.05$. Preliminary analysis assessing the effects of gender did not reveal any systematic differences. Thus, data were collapsed across gender.

Table 1 presents the means and standard deviations for percent of correct responses, as well as overall recall accuracy as a function of age, question format, and public vs. private response format. Separate ANOVAs were conducted for the percentage of correct answers and for overall accuracy, with age (6-, 7-, 8-year-olds) and response format (public vs. private) as between-subjects factors and question format (unbiased vs. misleading) as within-subjects factor.

An ANOVA on the percentage of correct answers revealed a significant main effect of age, $F(2, 116) = 10.45$. Subsequent Student–Newman–Keuls post-hoc tests indicated that the 8-year-olds (35.5%) outperformed both the 6- (19.4%) and the 7-year-olds (25%) with the two younger age groups not differing from each other. Importantly, the experimental manipulation (public vs. private response format) showed no main effect. There was no main effect of question format, but there was a significant interaction between question format and age, $F(2, 116) = 3.28$.

Table 1
Mean percentages of correct answers and mean overall recall accuracy as a function of age, question format, and public vs. private answers (standard deviations in parenthesis)

<table>
<thead>
<tr>
<th></th>
<th>Unbiased questions</th>
<th>Misleading questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private</td>
<td>Public</td>
</tr>
<tr>
<td><strong>6-year-olds</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Correct</td>
<td>18.1 (13.0)</td>
<td>24.2 (12.5)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>33.8 (16.5)</td>
<td>34.8 (31.2)</td>
</tr>
<tr>
<td><strong>7-year-olds</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Correct</td>
<td>24.0 (12.9)</td>
<td>21.2 (11.6)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>46.0 (27.3)</td>
<td>37.0 (22.2)</td>
</tr>
<tr>
<td><strong>8-year-olds</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Correct</td>
<td>29.6 (17.8)</td>
<td>29.7 (15.7)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>45.8 (25.4)</td>
<td>46.3 (20.1)</td>
</tr>
</tbody>
</table>
This was due to the fact that the 6-year-olds gave more correct answers to the unbiased questions (21.2%) than to the misleading questions (18%), while the 7- and the 8-year-olds gave more correct answers to the misleading questions (27.4% for the 7-year-olds, and 41.4% for the 8-year-olds) than to the unbiased questions (22.6% for the 7-year-olds and 29.7% for the 8-year-olds). There were no other interactions.

An ANOVA on overall recall accuracy [i.e., correct answers/(correct + incorrect answers)] revealed a significant main effect of age, $F(2, 116) = 8.28$. Subsequent Student–Newman–Keuls post-hoc tests indicated that the 6-year-olds (27.1%) were outperformed by the 7-year-olds (35.7%), and the 8-year-olds (44.8%) performed better than the 7-year-olds. There was also a main effect of question format, $F(1, 116) = 9.00$. Regardless of age, children showed higher accuracy in response to the unbiased questions (40.6%) compared to the misleading questions (31.3%). As was the case with the percentage of correct responses, there was again no main effect of private vs. public response format on recall accuracy, $F < 1$, and no interactions with question format.

### 2.4. Discussion

The aim of the present study was to investigate potential effects of private reporting on children’s recall accuracy and suggestibility in event recall interviews. Rather than using computerized interviews to provide an opportunity for private responses, we asked children in our private response condition to whisper their answers to unbiased and misleading questions into a teddy bear’s ear. In our public response condition, their responses to the same questions were given in a typical face-to-face interview situation. We hypothesized that the teddy bear would produce a more “private” recall situation than the standard interview, increasing children’s accuracy when responding to open-ended questions, and, more importantly, encouraging children to correctly reject the interviewer’s false suggestions.

The results showed no reduction in correct responses, incorrect responses or accuracy to unbiased questions, suggesting that the teddy was as effective as the human interviewer in eliciting accurate recall to unbiased questions. The analyses, however, failed to support our hypothesis regarding misleading questions. Regardless of age, children in the private and public response format answered similar percentages of misleading questions correctly, and additional analyses revealed that they also used the “don’t know” option with similar frequency. That is, children did not show any benefit from the opportunity to whisper answers to misleading questions into the teddy bear’s ear. There was also no suggestion that the private response condition encouraged them to offer answers about which they were uncertain and which they would have therefore withheld in the public response format in view of the accuracy instructions. Nor was there any evidence that the private response format encouraged them to correctly reject the interviewer’s false suggestion.

Since the manipulation check after the interview revealed that 73% of the children were absolutely sure that the teddy bear was discreet (and another 17% were
pretty sure about the teddy’s discretion), these results cannot be explained in terms of the failure of children to “buy” the private manipulation. In fact, reanalysis with exclusion of those children who were not absolutely sure of the teddy’s discretion produced no change in the pattern of results for any of the dependent measures. Further, we have no indication that the children did not take the task of whispering seriously; in fact, some children whispered so quietly that they had to be asked to speak up during the training phase (so that “teddy could hear”). There was also no indication that they did not try their best when reporting to the teddy, because there were no decrements in recall accuracy in the private response format.

This general pattern of no differences between private and public response condition for the misleading question format was unexpected given the large body of literature showing that children from an age of 5 years on are sensitive to the social demands of an interview situation and can adjust their answering behavior in accordance with these demands. However, studies reviewed in the introduction investigating the effects of interviewer-provided social support have sometimes also produced non-significant differences: a benefit of a warm atmosphere and positive communication has only yield to improvements in recall when contrasted with an intimidating, definitely non-supportive interviewer (Carter et al., 1996; Davis & Bottoms, 2002) but not when compared with a normal face-to-face interview (Goodman et al., 1991; Imhoff & Baker-Ward, 1999). Thus, one possible reason for the lack of significance differences in recall in our study and some of the social support studies may lie in the specific characteristics of the control conditions.

However, although there is evidence that social factors impact on suggestibility, the literature also clearly indicates an influence of cognitive factors, particularly on accuracy of event reports. Among the cognitive factors that have been shown to affect recall accuracy independently of memory, metacognitive monitoring processes, such as source attribution competencies and confidence judgments, play a dominant role. Moreover, metacognitive monitoring processes have been shown to be sensitive to manipulations of the social demands of the interview situation, in that children can perform metacognitive monitoring more adequately when social influences are reduced (Roebers, 2002; Roebers & Howie, 2003). This suggests the possibility that private reporting may improve older children’s ability to monitor the accuracy of their responses, rather than having its major impact on their recall accuracy per se. There is a large body of literature on the relation between monitoring (confidence) and accuracy of recall, suggesting that adequate monitoring does not necessarily lead to higher recall accuracy. (A review of that literature is beyond the scope of this paper; for reviews see Perfect, 2002.) Rather, the ability to reasonably act on the monitoring processes, that is, to engage in efficient metacognitive control processes, plays the crucial role. In fact, children’s metacognitive deficits seem to stem primarily from the translation of monitoring processes (e.g., “I am not sure that my answer just given is correct”) into control processes (e.g., “I’ll withhold my answer instead”; Schneider & Bjorklund, 1998). Ideally, of course, we would hope that efficient metacognitive monitoring of
retrieval will translate into more accurate recall, but if it can be established that metacognitive monitoring is superior with non-human compared with human interviewers under conditions where social demands are likely to operate, the focus of research and application could be on how to translate that better monitoring into better reporting. Thus, under normal conditions, monitoring processes can be independent or only weakly associated with recall and therefore deserve separate investigation. Our second study therefore focused on the effects of private versus public reporting on children’s monitoring of the accuracy of their recall responses.

This question is also of interest in relation to Roebers and Howie (2003) finding of poorer monitoring performance to misleading questions (in the form of inflated confidence in incorrect responses). Roebers and Howie raised the question of whether it is the actual process of metacognitive monitoring that is being undermined or contaminated by misleading questions, or only the reporting of that monitoring in the form of confidence judgments. It may be that children who publicly acquiesce to a misleading suggestion feel obliged to express confidence in their acquiescent response. If so, private reporting of metacognitive monitoring judgments should reduce any social pressures to exaggerate confidence in incorrect responses, thereby reducing the overconfidence effect, even though the level of acquiescence may not improve. A finding of no public-private difference in monitoring would suggest that misleading questions undermine the monitoring per se, not simply the reporting of that monitoring. In Study 2 therefore, we compared conditions in which reporting of confidence judgments was either private or public, although in both conditions the recall response itself was made publicly. This allowed us to ask whether it is at the level of metamemorial monitoring that privacy influences performance, rather than simply at the level of recall response given.

In Study 2, we explored the possibility that confidence judgments given privately to a computer would be more appropriate than those given publicly, as they have been in previous studies. In particular, we were interested in confidence judgments after misleading questions, because these appear to pose particular problems for young children. Because it was not practicable to use the Study 1 teddy bear procedure to obtain ratings on a 5 point scale, in Study 2 we adopted the procedure of computerized recording of ratings. We asked children the same questions about the event as in Study 1, and after answering each question, half of the children in each age group gave their confidence judgments openly and directly to the interviewer, while the other half privately keyed in their judgments on a computer keyboard, with the interviewer clearly unable to view these responses. Finally, in order to increase the contrast between open versus private metacognitive judgments, we manipulated social pressure in the recall situation by using question format as a between-subjects factor rather than as a within-subjects factor as in Study 1. Previous work has shown that children’s metacognitive monitoring is sensitive to the differences in the social demands produced by “bombardment” with questions which consistently follow the same format, as opposed to a mix of question formats, and this is particularly so when the bombardment question format is misleading (Roebers & Howie, 2003).
3. Study 2

3.1. Method

3.1.1. Overview

A $2 \times 2 \times 2$ factorial design was realized in this study. Two different age groups (8- and 10-year-olds) were asked either unbiased or misleading questions about an event and then gave either “private” or “public” confidence judgments after each answer with the help of a computer. That is, within each age group and question format condition, half of the children indicated their degree of confidence by clicking on the computer screen using a mouse, with the interviewer seeing, commenting on, and writing down the child’s confidence judgment on the protocol sheet. The other half of the children clicked on the computer screen in privacy, i.e., the interviewer patently had no knowledge of the child’s confidence judgments.

3.1.2. Sample

A total of 159 participants (54% girls) from two age groups (8- and 10-year-olds) completed the study. There were 79 children in grade 2 with a mean age of 8 years and 1 month (SD = 4 months). Forty of these 8-year-olds (18 girls and 22 boys) were randomly assigned to the unbiased condition, and 39 (19 girls and 20 boys) to the misleading condition. Within the unbiased condition, 20 (9 girls and 11 boys) were assigned to the private confidence judgments condition, while the remaining 20 (9 girls and 11 boys) were assigned to the public response format. Within the misleading condition, 20 8-year-olds (10 girls and 10 boys) were assigned to the private condition, while 19 (9 girls and 10 boys) were assigned to the public condition.

The 10-year-olds consisted of 80 children in grade 4 with an average age of 10 years and 2 months (SD = 24 months). Forty (24 girls and 16 boys) were assigned to the unbiased condition, and 40 (25 girls and 15 boys) to the misleading condition. Within the unbiased condition, 20 10-year-olds (12 girls and 8 boys) were assigned to the private confidence judgments condition, while 20 (12 girls and 8 boys) were assigned to the public condition. Within the misleading condition, 20 10-year-olds (13 girls and 7 boys) were assigned to the private condition, while the remaining 20 (12 girls and 8 boys) were assigned to the public condition.

The first part of the experiment, that is, the presentation of the video, was identical to Study 1. In the second part of the experiment, that is, the individual interviews, the instructions for the interview and the confidence judgments, as well as the training questions for the use of the confidence scale were the same as those used by Roebers and Howie (2003), except that the confidence judgments were delivered directly to the computer by clicking on the screen display rather than being reported to the interviewer, and results were automatically stored in a data base. The screen display for the confidence ratings depicted three “cards” (very sure, pretty sure, not sure) which were identical to those used by Roebers and Howie (2003). Prior to beginning the recall interview, children in all conditions were trained to give confidence judgments using the computer. This enabled the interviewer to check whether
the child was experienced in the use of the computer mouse. This proved to be the case for all children. As in Study 1, the delay between watching the video and the interview was on average 3 weeks.

In the public confidence judgments condition, children sat facing a portable computer with the interviewer seated beside them, also facing the screen so that it was clear to the child that she was aware of where and when the child clicked on the screen. This impression was confirmed by the fact that the interviewer also openly wrote down the child’s confidence judgments on the protocol sheet. This procedure also meant that after completion of the interview, it was possible to check for agreement between the protocol sheet and the computer input. There was 100 percent agreement for all protocols.

In the private confidence judgments condition, the interviewer sat across the table opposite the child and it was clear that she could not see which confidence judgment the child clicked on. In order to make the private confidence judgments condition more “private” and the instructions more credible, before the recall interviewer appeared, a confederate experimenter introduced herself and gave the instructions for the confidence judgments. The confederate explained that she wanted to find out how hard or easy the questions were and that she was not interested in the child’s memory for the event. She asked the child to click on the confidence judgments without letting the interviewer see. The reason given for this was that the interviewer believed that the difficulty of the questions was appropriate for kindergarten children, but the confederate wanted to make sure she was right, and to have some children’s opinions as well. In order to enable the interviewer to keep track of what the child was doing in the private confidence judgments condition without looking at the child’s responses, the computer was programmed to emit a sound for each mouse click. This enabled the interviewer to keep track of the child’s use of the computer.

3.2. Results

3.2.1. Recall

Although confidence judgments were the major focus of Study 2, we will briefly address the question of possible a priori recall differences between the two conditions in which the confidence judgments were given. Readers are reminded that in all conditions, children’s answers to the recall questions were given in public, that is, to the interviewer in a “usual” face-to-face interview situation. Only the confidence judgments were given in privacy for half of the children in each age group.

Table 2 presents overall recall accuracy as a function of age, question format and public vs. private confidence judgments. To test for a priori recall differences between the two experimental conditions, an ANOVA was conducted on overall recall accuracy [i.e., correct answers/(correct + incorrect answers)] with age, question format and public vs. private confidence judgments as between-subjects factors. This revealed a significant main effect of age that was due to the fact that 10-year-olds’ recall accuracy was higher [0.61] than 8-year-olds’ [0.55], $F(1, 151) = 4.76$. Importantly, there were no recall accuracy differences between the private and public confidence
judgment conditions, and no interactions between private vs. public conditions and any of the other main factors.

3.2.2. Confidence judgments

Table 3 presents the mean confidence judgments after correct and incorrect answers as a function of age, question format, and private vs. public confidence judgments. A MANOVA was conducted on the confidence judgments after correct and incorrect answers with correctness of answer to the recall questions as a within-subjects factor and age, question format, and public vs. private confidence judgments as between-subjects factors. The results revealed that confidence judgments were significantly higher after correct answers [2.5] than after incorrect answers [2.3], $F(1, 150) = 31.63$, significantly higher in 8-year-olds [2.5] than in 10-year-olds [2.3], $F(1, 150) = 9.27$, significantly higher after having answered a misleading question [2.5] than an unbiased question [2.4], $F(1, 150) = 6.86$, and significantly higher when given privately [2.5] than when given publicly [2.4], $F(1, 150) = 5.49$. There were also significant interactions between correctness of answer and age, $F(1, 150) = 19.17$, correctness of answer and question format, $F(1, 150) = 95.21$, and between age and privacy condition, $F(1, 150) = 3.92$. These interactions were further explored in separate ANOVAs of correct and incorrect responses.

An ANOVA on confidence judgments given after correct answers (age $\times$ question format $\times$ public vs. private confidence judgments) revealed that 8- and 10-year-olds did not differ from each other [2.5 for both age groups], $F(1, 151) = 0.05$, n.s., but
confidence judgments were significantly higher under unbiased questioning [2.6] than under misleading questioning [2.4], $F(1, 151) = 9.02$. Confidence judgments given privately about correct responses [2.6] were significantly higher than those given in public [2.4], $F(1, 151) = 4.82$. There was also a significant interaction between age and question format, $F(1, 151) = 4.24$, that was due to the fact that 8-year-olds’ confidence judgments given after correct answers to the unbiased questions [2.7] were significantly higher than the confidence judgments given after correct answers to the misleading questions [2.4], while the 10-year-olds’ confidence judgments after correct answers did not differ between the two question formats [2.5 and 2.5 for the unbiased and misleading question formats, respectively].

The corresponding ANOVA on the confidence judgments given after incorrect answers revealed that overall 8-year-olds’ confidence judgments [2.5] were significantly higher than 10-year-olds’ confidence judgments [2.2], $F(1, 151) = 25.48$, and that confidence judgments under misleading questioning [2.6] were significantly higher than under unbiased questioning [2.1], $F(1, 151) = 57.65$. Importantly, the public vs. private confidence judgments manipulation had no effect on the confidence judgments given after incorrect answers, $F(1, 151) = 2.78$, n.s.

3.3. Discussion

It is clear from the results that the private-public manipulation used in Study 2 did not affect the accuracy of children’s reported confidence judgments. Regardless of whether confidence judgments were made publicly, to the interviewer to whom they had given their recall response, or privately, to the computer and ostensibly for the benefit of a different adult who had distanced herself from the recall interviewer’s opinions, the pattern of results was strikingly similar to that reported by Roebers and Howie (2003) using public confidence judgments in a very similar context. Namely, both 8- and 10-year-olds showed appropriately differentiated judgments when questions were unbiased. On the other hand, when questions were misleading, the 8-year-olds showed overconfidence (greater confidence in incorrect than correct answers), and although the 10-year-olds showed no overconfidence, their confidence ratings still showed no ability to differentiate between correct and incorrect answers.

The opportunity to report privately did not overcome the difficulties faced by both age groups in making appropriate confidence judgments about their answers to misleading questions. This suggests that private, computerized interviews do not facilitate children’s metacognitive monitoring under suggestive or socially pressuring questioning. It also suggests that the undermining effect of misleading questioning on metacognitive monitoring performance noted by Roebers (2002) and Roebers and Howie (2003) does not simply reflect children’s reluctance to accurately report low confidence when acquiescing to misleading suggestion. Rather, and of greater concern, it appears to reflect actual undermining of the decision making/judgment processes.

It should be noted that our misleading question format condition involved “bombardment” rather than misleading questions intermixed with unbiased ones
in a within-subject manipulation. We would expect bombardment to maximize the social pressure, so if any effects of private reporting in reducing the impact of social pressure were to be found, we would expect to find them under these fairly extreme conditions. This, however, did not prove to be the case. It is possible that any potential for benefits from private reporting of confidence judgments was undermined by the impersonal nature of the computer procedure (clicking on a screen image using a mouse). As noted above, we were not able to overcome the practical difficulties of obtaining confidence judgments using the teddy procedure. It may be that more personal computer procedures, such as having a human or animated character speak on the computer monitor as in the Donohue et al. (1999) and Powell et al. (2002) studies, or having the child whisper a rating, might still prove to be beneficial.

Given that children’s confidence judgments are typically biased towards high confidence and that one aspect of developmental progression in confidence judgments concerns the overall level of confidence, favorable conditions for confidence judgments should yield first of all a decrease in the overall level of confidence. Contrary to our expectations, confidence judgments given into the computer were significantly higher than confidence judgments given directly to the interviewer. It is interesting that further analyses revealed that only the confidence judgments after correct responding showed this unpredicted effect, but this finding is difficult to interpret. However, it is possible that in the private condition children engaged in less cognitive testing of the veracity of their response because they were sure the interviewer would not be able to question it. If that was true, reaction times could serve as an indicator to confirm that interpretation. Unfortunately, we did not gather reaction times but further research could clarify this issue.

3.4. General discussion

Taken together, these two studies raise some intriguing questions about the role of computers and other forms of “private” reporting in investigative interviewing with children. In line with the findings of Donohue et al. (1999) and Powell et al. (2002) using recall prompts which were low in social demands, our results under conditions where social demands were higher show no beneficial effects of private reporting on recall or metacognitive monitoring, but, just as importantly, there are also no deleterious effects.

One reason why no benefits of private reporting were found in the present and the previous reported studies may be the influence of individual differences. Children may vary to a considerable degree in the way they respond to face-to-face interviews with adults. Temperamental and personality characteristics such as the approach/withdrawal dimension and shyness in social situations have been shown to be significantly associated with children’s event recall and suggestibility (Goodman & Quas, 1997; Gordon, Schroeder, Ornstein, & Baker-Ward, 1995; Quas et al., 1999; Roebers & Schneider, 2001b). These effects may mask any effects of manipulations of social demands. For example, children scoring high on the withdrawal dimension of temperament may be more likely to benefit from more private interview contexts than
children who enjoy being questioned by an unfamiliar adult. The considerable variability in most of the dependent measure in both of our studies is consistent with this argument.

The finding that private reporting failed to decrease suggestibility suggests that the problem of acquiescing to adults’ misleading questions derives not so much from embarrassment or from reluctance to contradict adults to their face, but rather from a simple conviction that if an adult thinks that something happened, then children believe it most probably did. This argument relates to the concept of normative versus informative social influences (Deutsch & Gerard, 1955) and developmental differences in their impact (Schwarz et al., 2004). Because children encode information in a less elaborated and less complete form than adults (Marche & Howe, 1995) the informative aspect of any social influence is likely to be very strong, and stronger than any normative social influences. In our study and in previous studies, the procedures used to decrease social influences during interrogation involved systematic manipulation of the normative influences while the possibly more important informative social influences remained relatively untouched. This is one possible reason why the social manipulations in our and the previous studies were ineffective. Future work needs to consider more carefully the different sources and aspects of social influences when manipulating social demands of interview situations.

Looking at previous experimental manipulations of the social factors in children’s investigative interviews, one could argue that significant positive effects have been found only in those studies in which the favourable interview condition was contrasted with rather extreme interview conditions of the sort that would be considered to be highly negative and inappropriate in real life (Carter et al., 1996; Davis & Bottoms, 2002). In studies in which the control condition was a more “normal” interview situation, potential beneficial effects of, for example, interviewer-provided emotional support are absent or fail to reach significance. That is, the private-public dimension may still prove to be a critical one, but only if there is a very pronounced contrast between public and private conditions. An alternative possibility which would be worth pursuing empirically is that less pronounced contrasts still yield significant differences in suggestibility or in recall accuracy, but only when the events being recalled are highly sensitive, personally relevant or embarrassing.

In spite of the largely negative results of our studies, it is still possible that modest variations in contextual and other factors may prove to facilitate children’s event recall. At this early stage of research on the impact of social demands affecting children’s event recall and suggestibility, the methodologies of the existing empirical studies are very heterogeneous, especially with respect to the procedures for social manipulations. Our two studies, and those in the previous literature, differ on a range of dimensions which need to be systematically tested. Specifically, the relevance of the “persona” represented on the computer screen needs to be systematically examined. Does the representation of the face of an adult, human authority figure produce different performance from the representation of a fantasy or cartoon character, such as Marvin the Mouse as used by Powell, Wilson, and Hasty (2002)
to evaluate children’s linguistic and conceptual abilities, or a disembodied human voice, or written questions appearing on the screen? What is the impact of whispering responses to an inanimate toy as opposed to whispering it to a computer? If benefits of private reporting are found, is it the privacy of the responding that is the key ingredient, or the absence of face-to-face confrontation with another human being, or simply the novelty of talking to a non-human? How might these procedures be affected by children’s shared understanding of the serious purpose of the interview in forensic contexts and their belief in the discretion of the non-human recipient of their responses? What is the effect of the strength of the child’s memory and/or the delay since experiencing the reported event? How do all these factors apply to children of different ages? The answers to these questions have major implications for the application of these kinds of procedures in investigative interview settings with children.

On the other hand, it is important to note that findings to date on the impact of children’s private recall and monitoring performance are consistent in that they all fail to demonstrate significant benefits of private reporting in comparison to face-to-face interrogations. This seems especially noteworthy because a variety of commonly used measures of children’s eyewitness memory and suggestibility have been used. Moreover, the studies produced a consistent pattern of results despite the fact that the nature of the private recipient was different across studies, despite different to-be-remembered or disclosed information, and despite significant differences in the delays before recall and thus the strength of children’s memory. Taken together, our results and the previously reported findings may be interpreted as showing that the impact of social factors has been somewhat overestimated.

Acknowledgements

The present studies are associated to two larger research projects on the influence of social and metacognitive factors on children’s event recall financed by the German Research Foundation (DFG-Gz. FOR 261/2-3 and DFG-Gz. RO 1324/4-1). We wish to thank Melanie König and Katharina Seidler for their help with the data collection as well as the children, teachers, principals and parents for their cooperation and participation.

References


