# Chapter 8
# The Realism in Children's Metacognitive Judgments of Their Episodic Memory Performance

**Carl Martin Allwood**

## 1 Introduction

This chapter deals with the metacognitive realism in children's confidence judgments of their own episodic memory performance. *Metacognitive realism* is sometimes called metacognitive accuracy and is defined as the extent to which a metacognitive judgment is veridical with respect to the asserted status of the learning or of the recall (Lichtenstein, Fischhoff, & Phillips, 1982). After some remarks on metacognition, the realism in metacognition and its measurement, the chapter focuses on some factors that can influence the realism in confidence judgments of the correctness of one's episodic recall. Two studies are then presented more in detail in order to deepen the analysis of some of the factors envisaged. The chapter also relates the level of realism in children's confidence judgments to that of adults.

Metacognition is usually described as our knowledge about our own cognition, including the use of this knowledge to regulate our own cognitive processes (Weinert & Kluwe, 1997). Sometimes it is also taken to include knowledge about other people's cognition (Allwood & Granhag, 1999; Allwood & Johansson, 2004; Jost, Kruglanski, & Nelson, 1998). The approach on metacognition taken in this chapter, just as in current research (Koriat, 2007), sees metacognition in a system perspective where metacognitive processes are integrated into the individual's other cognitive processes and where they are also affected by various social processes taking place outside the individual. For example, retelling the event to different persons after the experience and taking part of one's listeners' reactions to one's story is likely to influence both the correctness of later recalls of the event and one's confidence that the recall is correct.

C.M. Allwood (✉)
Department of Psychology, University of Gothenburg, Box 500,
SE-40530 Gothenburg, Sweden
e-mail: cma@psy.gu.se

A. Efklides and P. Misailidi (eds.), *Trends and Prospects in Metacognition Research*,
DOI 10.1007/978-1-4419-6546-2_8, 

### *1.1 Realism in Confidence Judgments*

Confidence judgments are judgments of the veridicality of one's own or other's memory reports. At a general level the *realism* in confidence judgments is a function of the relation between the correctness of performance (as measured against a socially approved standard) and the performer's confidence in the correctness of his or her performance. As pointed out by Leippe (1980), although there are factors that may influence both the correctness of the memory report and the person's confidence that the memory report is correct (such as good performance conditions in general), some factors may predominantly influence the level of correctness whereas others may foremost influence the level of confidence. For example, an individual's confidence level may partly be a function of the individual's general level of expression of confidence over time; that is, individuals tend to have a certain stability in the level of confidence they express (see, e.g., Jonsson & Allwood, 2003).

There are at least two aspects to the realism of metacognitive judgments (for a much more complete presentation of these issues, see Yates, 1994). The two aspects can be distinguished through the use of many measurements for an individual or a group of individuals. These are *calibration* and *discrimination*. The calibration aspect pertains to the relation between the *level of confidence* in the correctness of the memory recall and the *level of correctness* in the same memory recall. When these two levels coincide the judge is said to show perfect calibration. The other aspect, discrimination, pertains to the individual's ability to discriminate between *correct* and *incorrect* items by means of his or her confidence judgments. Each of these two aspects can be measured in different ways.

Appendix 8.1 shows some common measures used in the calibration tradition branch of metacognitive research. The measures shown are calibration, over-/underconfidence, resolution, and slope. The first two of these measures relate to the calibration aspect. Of the last two, resolution relates to the discrimination aspect. Slope measures the separation between confidence for correct and incorrect items and has an advantage in that it may be more intuitively easy to understand than resolution. It picks up discrimination ability but also some of the calibration aspect.

## 2 Factors Influencing the Realism in Confidence Judgments

A host of different factors can influence metacognitive realism in confidence judgments (see, e.g., Allwood & Granhag, 1999; Jonsson & Allwood, 2003; Klayman, Soll, Gonzáles-Vallejo, & Barlas, 1999). Table 8.1 presents some factors that can be assumed to influence the realism in confidence judgments in episodic memory recall.

Although not mentioned in Table 8.1, these factors do not preclude the influence of cognitive processing biases (Koriat, Lichtenstein, & Fischhoff, 1980) such as the confirmation bias, and methodological and statistical factors (Erev, Wallsten, & Budescu, 1994; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Juslin, 1994) such as

**Table 8.1** Factors that may influence the realism in children's confidence judgments of their episodic performance

| Event |
|---|
| The individual's general level of expressed confidence |
| Intermediate events (communication with others, etc.) |
| Social situation in which memories are reported and confidence judgments are made |
| Memory question asked (type of question, content in question, e.g., central/peripheral aspect) |
| Measurement scale |
| Aspect of realism in confidence analysed (e.g., calibration or discrimination) |

biased selection of items, on confidence judgments of episodic memory reports. Various cognitive processing biases and methodological and statistical factors have received researchers' attention, especially in the context of semantic memory and for adults, but are less well researched for episodic memory and children. However, they fall outside of the focus of the present chapter.

Initially it can be noted that various features of the to-be-remembered event and its context, such as visibility and other encoding conditions are likely to influence the correctness of the memory recall and thus also the realism of the confidence judgments (see, e.g., Leippe & Eisenstadt, 2007). However, this factor (Factor 1 in Table 8.1) is not elaborated in the present chapter. Furthermore, as noted above, individual stability in the general level of confidence judgments (Factor 2 in Table 8.1) as such may also influence the realism of confidence judgments but this also falls outside the scope of the present chapter.

## *2.1 Events Intervening Between the Original Event and the Memory Report*

Different *events intervening* in the time between the original event and when the memory is reported and confidence judged may influence the realism of the confidence judgments. Importantly, as noted above, social conversations of various types, that is, talk with other persons, could have an influence. For example, when the child reports and discusses the event with his/her family, friends and other categories of persons such as the police or other parties in a forensic process, various memory consequences are likely to take place. One important consequence is that the memory of the event is actively rehearsed verbally, and this is likely to increase the correctness in future recall of the memory (Roediger & Karpicke, 2006).

However, given that the event is not just reported but also *discussed*, both correctness and confidence may be influenced (see, e.g., Marsh, 2007). Correctness might be influenced, for example, due to the fact that the conversation partners ask questions that express or imply erroneous assumptions that are then encoded in the person's memory in such a way that these assumptions or their implications cannot be distinguished from the original event.

In addition, previous research shows that confidence in the correctness of a statement may increase as an effect of reasserting it. This has been called the *reiteration effect* (Hertwig, Gigerenzer, & Hoffrage, 1997). Thus, each time a memory is reasserted, for example when retelling the event or answering questions about it, the child's confidence in the correctness of the memory report may increase.

The effect of intermediate discussions of the event on the rememberer's correctness and confidence was investigated in students by Sarwar, Allwood, and Innes-Ker (2010a). In that study repeated retellings of the events in the film (approaching simple *repetition* of the experienced event) were found to increase correctness, confidence and the realism (as measured by the calibration measure, see Appendix 8.1) in the confidence judgments of the correctness of the reported memories in the context of a later open free recall task. In contrast, multiple *discussions* of the event with (each time a new) other person reduced the effects of repetition in the later free recall task, that is, compared with a control condition, discussions did not significantly affect any of the mentioned measures. To these authors' knowledge no similar study has been made for children.

## *2.2 Social Aspects of the Memory Report Situation*

As shown in Table 8.1, *the type of social situation* in which the child recalls and then confidence judges the correctness of the recalled information might also influence the realism of the confidence judgments. How the child experiences and understands the situation is important here. For example, situations vary with respect to the extent to which the child feels expected to only report correct information from memory. In fact, the child may be given an explicit instruction to only provide information that they are absolutely sure is correct, or, alternatively, to report anything that might be true even if they are not sure about it. For example, when testifying in court a child can in ordinary situations be assumed to attempt to only provide correct information and for this reason choose not to report memories about which they feel unsure about. On other occasions the child might apply a less stringent threshold for reporting memories; for example, in free time discussions when they want to impress their friends.

Koriat and Goldsmith (1996) presented a model for how confidence judgments are integrated in ordinary memory recall. The point of the model that is relevant in the present context is that it assumes that the rememberer uses confidence judgments to regulate which memories are reported. This is accomplished by the implementation of a variable threshold for how sure he or she wants to be that the reported memories are correct. Accordingly, when a person can choose what memories to report (i.e., using Koriat and Goldsmith's (1996) term, they have *free report option*) they can themselves attempt to regulate the assumed proportion of correct memories.

For example, Koriat, Goldsmith, Schneider, and Nakash-Dura (2001) analysed 7–12 year-old children's answers to specific questions with and without answer

alternatives about a slide show that they had seen. The results showed that when the child could control what information to report, correctness tended to increase, and completeness decrease compared with when they had to/were pressed to answer. To sum up, various constraints in the social situation where the recall and confidence judgments take place can affect the level of both correctness and confidence.

## *2.3 The Memory Question Asked*

The factors presented in Table 8.1 also bring attention to the importance of the memory question asked for the ensuing realism of the confidence judgments. For example, the degree of veracity of the information provided in the memory question is important. Previous research has shown that children have difficulties with misleading questions (Roebers, 2002; Roebers & Howie, 2003). For example, Roebers and Howie (2003) studied 8- and 10-year-old children's and adults' discrimination of correct and incorrect answers to memory questions on specific aspects of a short film clip by means of their confidence judgments. In this research unbiased and misleading questions (questions that suggested an incorrect answer) were compared. Misleading questions were more difficult for the children to handle. For *unbiased questions* all three age groups gave higher confidence judgments for correct answers than for incorrect answers. However, for *misleading questions* only adults gave higher confidence judgments for correct answers than for incorrect answers.

There are also a number of other important aspects to the question asked. First, the question asked can vary with respect to *how much information* it provides about the to-be-reported memory (for example, free recall or recognition questions). Everything else being equal, recognition questions usually promote higher correctness than open specific questions since they provide more cues to the answer.

Second, questions differ with respect to how broad the assigned *answer-area* is that the question allows. Everything else being equal, the broader the assigned answer-area the more report control the remembering person has. For example, in open free recall questions the person is given a general indication about which area to report on ("Tell me all that happened on the Monday afternoon…") and within this area it is up to the person's own discretion what exactly to report. For more specific questions such as "What was the colour of the girl's jacket?" the assigned answer-area is much smaller.

Third, the *giving of report option* can as such, at least partly, be manipulated independently of how much information is provided about the to-be-reported memory and how large the assigned answer-area is. For example, for a set of recognition questions, the person may be told, or not, that he or she can choose which of the questions he/she wants to answer.

A fourth and final aspect of the memory question that can affect the realism in confidence judgments is the *type of contents* asked for. For example, Sarwar et al. (2010b) found indications that central information, such as the culprit's features or actions, may not only be better remembered but also more realistically confidence

judged than peripheral information, such as details not immediately relevant to the central action or to the actors. This may to a large extent be a function of how the person's attention is allocated during the encoding of the original event and of the person's prior knowledge about the contents (see, e.g., Christianson & Loftus, 1991; Ibabe & Sporer, 2004).

### *2.4 The Measurement Scale Used and the Aspect of Metacognitive Realism Analyzed*

As shown in Table 8.1, it has also been suggested that the specific confidence scale used to measure children's confidence may affect their ability to give realistic confidence judgments (Roebers & Howie, 2003). This suggestion is supported by the findings reported by Tunney and Shanks (2003), that is, students who used a binary confidence scale showed better realism in their confidence judgments of fairly implicit knowledge compared with participants who used a continuous scale.

A number of different confidence scales have been used in previous research to measure children's confidence (e.g., Allwood, Granhag & Jonsson, 2006a; Dirkzwanger, 1996; Newman & Wick, 1987; Roebers, 2002). One reason for this variation is the controversy concerning how complex tasks and scales children can handle at different ages. For example, numerical scales might be more (or even, too) complex for younger children compared with scales using qualitative steps, such as "Very unsure", "Not so sure", "Neither unsure nor sure" "Pretty sure" and "Very sure" (used by, e.g., Roebers, 2002).

Roebers and Howie (2003) suggested that one reason why younger children (e.g., 8-year olds) might perform poorly on metacognitive tasks could be that they are tested with age-inappropriate scales, specifically, scales that are too complex. Other inappropriate features of scales, such as use of *smiley faces* with broader smiles for higher confidence levels were also remarked upon by these authors. Roebers (2007) suggested that scales with fewer scale steps, for example three, would be easier for young children to handle. (The issue of scale inappropriateness is further discussed below, in Sects. 3 and 6.)

Finally, the factors presented in Table 8.1 suggest that the aspect of metacognitive realism analyzed (e.g., the calibration or the discrimination aspect) is likely to influence the conclusions drawn about the level of realism in children's metacognitive performance. For example, as discussed below in Sect. 5, children and adults may differ for some tasks in the calibration aspect but not in the discrimination aspect.

Next, two studies are reviewed that investigated the importance of two of the factors presented in Table 8.1 for the resulting metacognitive realism. These factors are the measurement scale used when the participants give their confidence ratings and the memory question asked.

## 3 The Effect of the Confidence Scale Used on Children's Metacognitive Realism

Allwood, Granhag, and Jonsson (2006a) analyzed the effect of the confidence scale used to measure the children's confidence in their memories of an experienced event. The participants were 81 children (41 girls and 40 boys) aged 11–12 years from Grades 5 to 6, in schools located in a middle class area in southern Sweden.

Four confidence scales were investigated: (a) Numeric, (b) Picture, (c) Line, and (d) Verbal scale. These scales are shown in Appendix 8.2. The rationale for including the respective scales in the study was as follows. The Numeric scale is common in calibration research with adult participants and it was included in order to allow for comparison with results from such studies. Picture scales (smilies, etc.) are common in research with younger children (Roebers, 2002). A Picture scale was included in order to be able to compare the results from this scale with those from the Numeric scale. The use of the Line scale was inspired by results reported by Nilsson (1998, p. 97), who concluded that younger children's (6-year-olds and to some extent 10-year-olds) handling of probabilities might be influenced by "perceptual factors such as size, shape and colour." The Line scale was included in order to examine if the participants' confidence ratings would be improved by a scale that highlights spatial aspects. Finally, Teigen (2001) presented results for adults that showed that written probability phrases tended to differ from numerical probabilities. Given this, we included the Verbal scale since we wanted to examine if the inclusion of written probability statements would influence the level of children's confidence judgments. Based on a review of earlier research we predicted that no difference would be found between the scales with respect to the level of the confidence ratings, nor with respect to the realism in the confidence judgments.

### *3.1 Method*

A between-subjects design was used, that is, the participants were randomly divided into four conditions (i.e., *numeric*, *picture*, *line* and *verbal*) and in each condition one of the four respective confidence scales was used.

#### 3.1.1 Procedure

The participants first watched a videoclip (approximately 4 min long) showing the kidnapping of a woman by two men who pulled her into a car by force. After viewing the videotape, the participants were given a 10 min training session on probability assessments. In this training the participants were provided with general explanations about probability estimates. For example, it was explained that a scale value of 60% meant that in the long run 60% of the items they had confidence rated as

**Table 8.2** Means (and *SD*) of correctness, confidence, calibration, over-/underconfidence, and resolution for the four scales (numeric, picture, line, and verbal)

| | Scale | | | |
|---|---|---|---|---|
| | Numeric | Picture | Line | Verbal |
| | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) |
| Correctness | 0.59 (0.09) | 0.58 (0.08) | 0.56 (0.08) | 0.57 (0.08) |
| Confidence | 0.81 (0.09) | 0.82 (0.08) | 0.76 (0.10) | 0.78 (0.07) |
| Calibration | 0.10 (0.06) | 0.10 (0.05) | 0.09 (0.05) | 0.09 (0.05) |
| Over-/underconfidence | 0.22 (0.13) | 0.24 (0.11) | 0.20 (0.12) | 0.22 (0.09) |
| Resolution | 0.04 (0.02) | 0.03 (0.02) | 0.03 (0.02) | 0.03 (0.02) |

60% sure should be correct. In addition, the children solved practical examples, guided by the experimenter.

Next, each participant answered 44 two-alternative directed questions on various details in the video (forced choice). After each question, the participants immediately rated their confidence in the correctness of the answer on a rating scale ranging from 50% (guessing) to 100% (completely sure).

### *3.2 Results*

The results for the various dependent measures are shown in Table 8.2 for each of the four confidence scales. Since the random chance for selecting the correct answer was 50%, the results for correctness (56–59%) shows that the questions were rather difficult. As expected, the results showed no significant differences between the four scales for correctness or for confidence.

The calibration aspect of realism in metacognition was shown by the specific measures calibration and over-/underconfidence. The discrimination aspect was shown by means of the resolution measure (see Appendix 8.1). Again, as expected, neither of these measures showed any significant differences between the four confidence scales.

## 4 The Effect of Question Type on Children's Metacognitive Realism

The study by Allwood, Innes-Ker, Homgren, and Fredin (2008) analyzed the effect on children's metacognitive realism of asking open free recall (henceforth called *free recall*) and specific directed questions (henceforth called *focused questions*) about an experienced event. We also analyzed the effect of repetition of answers on the realism in confidence judgments. We expected that both the children and the adults would show equal, and good, levels of metacognitive realism with respect to

the over-/underconfidence measure for their free recall. We also expected poor metacognitive realism in the confidence judgments of both children and adults of the answers to the focused questions.

## *4.1 Method*

The same procedure was used in two experiments. The only difference was that a different film clip was used as the to-be-remembered event. Information about the participants is shown in Table 8.3.

### 4.1.1 Procedure for Experiment 1 and 2

Each experiment had three sessions. In Session 1 the participants first saw a short 3–4 min videotape. In Experiment 1 the videotape was about a man who was looking for his lost dog in a park. The video in Experiment 2 was the same as the video used by Allwood, Granhag et al. (2006a) showing a kidnapping event. This video clip is likely to have been more complex than the video used in Experiment 1.

In Session 2, 1 week after Session 1, each participant was tested individually and all interviews in this session were audio-taped. First, the participants were asked to give a *free recall* of the events in the video clip they had watched 1 week before. The participant was asked to tell everything he or she could remember about the events in the video and the people in it, with as much detail as possible. Next the participant answered a questionnaire with 39 (Experiment 1) or 44 (Experiment 2) two-alternative forced-choice questions on specific details in the video (i.e., the focused questions). Two examples of questions used in Experiment 1 are "Was there a sandbox in the park shown in the film? Answer: (a) Yes, (b) No" and "How many swings were there in the park? Answer: (a) 3, (b) 5".

During the week between Session 2 and 3, the free recall data were segmented into elementary statements and further prepared for the confidence judgments in Session 3. When doing this segmentation the researchers followed the guidelines presented in Allwood, Ask, and Granhag (2005). A numeric confidence rating scale was placed under each elementary statement. The participants' answers to the focused questions were similarly prepared for the confidence rating in Session 3. For each question, a numeric confidence rating scale was placed immediately under the two answer alternatives (whereof the participant had selected one).

**Table 8.3** Participants in Experiment 1 and 2 in the Allwood et al. (2008) study

| 8–9-year-olds | 12–13-year-olds | Adults |
|---|---|---|
| *Experiment 1* | | |
| 31 (20 girls) | 31 (15 girls) | 32 (21 women); $M = 25$ years, Range = 19–56 years |
| *Experiment 2* | | |
| 43 (20 girls) | 52 (24 girls) | 38 (25 women); $M = 25$ years, Range = 18–46 years |

Session 3 took place in the children's classrooms. First, the children were given a 10 min explanation about confidence ratings. This explanation included specific examples and a detailed explanation about what a confidence judgment is. Next, the participants first confidence rated each item in the questionnaire with the statements in their free recall and then their answers to the focused questions. The confidence rating scale for both types of ratings went from 0 (I am sure that the answer is wrong), via 50 (I guess) to 100 (I am sure that the answer is correct). The same procedure but with suitable adjustments, was used for the adults; for example, the explanation about confidence ratings was shorter.

## *4.2 Results Experiment 1*

Table 8.4 shows some important results for the free recall task and for the focused questions task in Experiment 1. The results are reported in more detail in Allwood et al. (2008). The last column in Table 8.4 shows the outcome of planned contrasts between the three age groups. We first consider the results for the *free recall*.

First, as shown in Table 8.4, the analysis of the total number of statements recalled showed that the 8–9-year-olds and the 12–13-year-olds on average recalled significantly less statements than the adults. From youngest to oldest, the three age groups recalled 12.1, 12.3, and 15.6 statements, respectively. Moreover, both the correctness and the confidence levels were quite high. From 74 to 84% of the statements in each age group were located at the 100% confidence level and over 90% of these were correct. No age differences were found for correctness and the 12–13-year-olds tended to show the highest confidence.

We next look at the metacognitive measures. Here it is most noteworthy that the youngest age group had close to perfect realism in the over-/underconfidence measure.

**Table 8.4** Experiment 1: Means (*SD*) and number of participants [*n*] in the three age groups for the free recall and for the focused questions

| | 8–9-year-olds | 12–13-year-olds | Adults | |
|---|---|---|---|---|
| | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *p* |
| *Free recall* | | | | |
| Number of statements | 12.1 (4.3) | 12.3 (4.6) | 15.6 (3.9) | 1, 2,<3, $p<0.05$ |
| Correctness | 91.0% (10.9%) | 88.3% (10.5%) | 87.8% (11.1%) | *ns* |
| Confidence | 90.2% (9.5%) | 94.4% (6.1%) | 92.1% (5.9%) | $p=0.08$ |
| Over-/underconfidence | −0.005 (0.093) | 0.065 (0.100) | 0.042 (0.101) | 1<2, $p<0.05$ |
| Slope [*n*] | 13.5 (21.7) [19] | 10.5 (21.1) [23] | 3.7 (22.3) [26] | *ns* |
| *Focused questions* | | | | |
| Correctness | 71.7% (6.7%) | 74.8% (10.5%) | 73.5% (6.6%) | *ns* |
| Confidence | 78.4% (13.0%) | 85.0% (8.5%) | 74.1% (8.7%) | 1<2>3, $p<0.05$ |
| Over-/underconfidence | 0.068 (0.131) | 0.101 (0.110) | 0.008 (0.085) | 1, 2,>3, $p<0.05$ |
| Slope [*n*] | 7.8 (11.4) [31] | 13.9 (10.6) [31] | 12.8 (6.4) [32] | 1, 2,>3, $p<0.05$ |

*Note*: 1 = 8–9-year-olds; 2 = 12–13-year-olds; 3 = Adults

However, the other age groups also showed quite good values. *Slope* was used as a measure of ability to use confidence judgments to separate correct from incorrect answers (see Appendix 8.1). However, as shown in sharp brackets in Table 8.4, only 19 (61%), 23 (74%), and 26 (81%) of the participants in each age group (starting with the youngest) had any errors at all. Still, it can be noted that no age differences were found.

The results for the *focused questions* are also shown in Table 8.4. Here the confidence judgments were much more evenly spread over the confidence scale, but with concentrations at the 50% and the 100% level. No age differences were found for correctness and the 12–13-year-olds showed the highest confidence. For the metacognitive measures it is noteworthy that the adults demonstrated next to perfect realism with respect to the degree of overconfidence in contrast to the two child groups who showed overconfidence. The youngest age groups showed poorer slope than the other two groups.

It can be noted that the 12–13-year-olds and the adults showed less overconfidence for the focused questions than similar groups in our previous research (e.g., Allwood, Granhag, et al., 2006a; Allwood, Granhag, & Johansson, 2003). Two possible reasons for this are, first, that the contents of the film had a rather simple and monotonous structure which may have made the task simpler. Second, *repetition* of some assertions in the focused questions that were already made in the free recall may have caused an increase in confidence for these items due to the reiteration effect (Hertwig et al., 1997). (Seen from a forensic point of view, the non-overlapping items, that is, the answers to the focused questions not already mentioned in the free recall, are the most interesting since they may provide new information compared with the witnesses' free recall.)

To study the effect of repetition in the focused questions, the data for the focused questions was analyzed again. This time items already mentioned by a participant in the free recall were not included for that participant. Over all participants, 14.2% of all focused questions overlapped with some content mentioned in the free recall (522 questions out of 3,666 questions). The interjudge reliability for this coding was 88% and there were no age differences in number of overlapping items.

Table 8.5 shows the results for the focused questions when the focused questions that overlapped with content mentioned in the free recall are excluded. As can be

**Table 8.5** Experiment 1: Means (*SD*) and number of participants [*n*] in the three age groups for the focused questions when the questions overlapping between the free recall and the focused questions are excluded

| | 8–9-year-olds | 12–13-year-olds | Adults | |
|---|---|---|---|---|
| | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *p* |
| Correctness | 58.2% (7.1%) | 62.1% (9.1%) | 59.4% (7.2%) | *ns* |
| Confidence | 76.5% (12.8%) | 83.3% (9.4%) | 71.6% (9.0%) | 1<2>3, $p=0.05$ |
| Over-/underconfidence | 0.185 (0.122) | 0.215 (0.130) | 0.123 (0.099) | 1, 2>3, $p<0.05$ |
| Slope [*n*] | 1.3 (9.4) [31] | 5.1 (10.2) [31] | 1.9 (10.2) [32] | *ns* |

*Note*: 1 = 8–9-year-olds; 2 = 12–13-year-olds; 3 = Adults

noted by comparing Tables 8.4 and 8.5 one effect of excluding overlapping questions was that the level of correctly answered questions decreased by about 13%. However, the level of the confidence judgments was fairly unaffected and thus, there was no clear sign of a reiteration effect in this fairly realistic test of this effect. Due to the decrease in correctness the level of overconfidence increased substantially by about 0.12. Finally, the slope measure decreased to nearly zero for all groups. These results also show that when the overlapping items are not included in the results for the focused questions there was a clear format difference between the free recall task and the focused questions task.

## *4.3 Results Experiment 2*

First, the analysis of the total number of statements recalled showed that the 8–9-year-olds on average recalled significantly less statements (6.4) than both the 12–13-year-olds (9.8) and the adults (17.6). The latter two groups also differed in the number of recalled statements.

Table 8.6 shows that correctness was again high in the *free recall* task. For the 8–9-year-olds and the adults it was over 90%. Again a high percentage of the items (76–87%) were located at the 100% confidence level and over 90% of these were correct, except for the 12–13-year-olds who scored about 85%. Furthermore, the results showed that the 12–13-year-olds had significantly poorer correctness than the other two age groups, but still high at 84%.

For the metacognitive measures it is again striking that the youngest age group had close to perfect realism for the over-/underconfidence measure, but the adults

**Table 8.6** Experiment 2: Means (*SD*) and number of participants [*n*] in the three age groups for the free recall and for the focused questions

| | 8–9-year-olds | 12–13-year-olds | Adults | |
|---|---|---|---|---|
| | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *p* |
| *Free recall* | | | | |
| Number of statements | 6.4 (3.6) | 9.8 (3.8) | 17.6 (9.9) | 1, 2<3, 1<2 $p<0.05$ |
| Correctness | 92.5% (15.0%) | 83.8% (12.4%) | 90.7% (8.2%) | 1>2, 2<3 $p<0.05$ |
| Confidence | 91.4% (13.1%) | 93.4% (7.5%) | 93.4% (7.5%) | *ns* |
| Over-/underconfidence | –0.004 (0.158) | 0.096 (0.116) | 0.034 (0.084) | 1<2, 2>3, $p<0.05$ |
| Slope [*n*] | 11.7 (34.5) [12] | 10.5 (20.5) [44] | 28.3 (36.8) [28] | 2<3, $p<0.05$ |
| *Focused questions* | | | | |
| Correctness | 54.4% (8.7%) | 56.7% (6.3%) | 54.4% (7.0%) | *ns* |
| Confidence | 73.5% (13.0%) | 68.9% (10.1%) | 72.0% (9.4%) | *ns* |
| Over-/underconfidence | 0.194 (0.169) | 0.122 (0.109) | 0.176 (0.117) | 1>2, $p<0.05$ |
| Slope [*n*] | 1.5 (7.7) [43] | 4.1 (5.7) [52] | 3.2 (7.2) [38] | *ns* |

*Note*: 1 = 8–9-year-olds; 2 = 12–13-year-olds; 3 = Adults

also showed quite good realism. With respect to slope only 28, 84, and 74% of the participants in each age group (starting with the youngest) had any errors and thus were available for the analysis. In this experiment the adults had better slope than, especially, the 12–13-year-olds.

For the *focused questions* it should be noted that the adults showed higher overconfidence than in previous research (unclear why). This deviation from previous research can not be explained by the repetition of items that took place between the free recall and the focused questions since an analysis showed that over all participants, only 1.7% of all focused questions overlapped with some content mentioned in the free recall (103 of 5,779 items). This coding had 91% interjudge reliability and ANOVAs showed no difference from the results compared with when all items were included for the focused questions.

## 5 Comparison of the Realism in Children's and Adults' Confidence Judgments

Metacognitive ability is usually assumed to improve as children get older. For this reason it is of interest to see what empirical research shows with respect to this issue in the domain of confidence judgments of event memory. Most previous studies on the realism in children's confidence judgments have used some form of specific directed questions. Next, we will compare the results from children with the results from adults (students) in previous studies using the same film clip and the same confidence scale, and when the comparison concerns focused questions, also the same questions.

### *5.1 Overconfidence*

The value for overconfidence for adults (students) in the relevant conditions in Allwood, Granhag, and Johansson (2003), Allwood, Knutsson, and Granhag (2006b), Granhag (1997) and Granhag, Strömwall, and Allwood (2000) varied between $M = 0.061$ and $M = 0.127$. In contrast, the overconfidence for the children in Allwood, Granhag, et al. (2006a) was on average 0.22 (see Table 8.2).

The study reviewed above by Allwood et al. (2008) is not strictly comparable with these studies for the focused questions since these questions in Allwood et al. (2008) were preceded by a free recall session. However, the results for the free recall in Experiment 2 in that study can be used for comparison since the same film was used. As reviewed above, it was shown that there was no significant difference between the children and the adults (the youngest children showed next to perfect realism for this measure).

These results suggest that before conclusions about developmental age differences are drawn in metacognitive research it is important to consider how well practiced the participants are at the specific task they are asked to perform. When the task can

be assumed to be well-trained (for example regulating the veracity of the output from memory in response to a free recall instruction), children can be expected to perform well, often at the level of adults.

### *5.2 Discrimination and Separation Measures*

Most of the above mentioned previous studies have also included measures of discrimination. Here a comparison of the results for children and adults show no great differences for the *resolution* measure (a measure of discrimination ability) when *focused questions* were used. The results generally show a resolution level of between 0.03 and 0.04 for both groups (Allwood et al., 2003; Allwood, Granhag, et al., 2006a; Allwood, Jonsson, & Granhag, 2005b; Allwood, Knutsson, et al., 2006b; Granhag, 1997).

The study by Allwood et al. (2008) used another measure, *slope* that in addition to separation as given in terms of scale values, also indicates discrimination. These results are less reliable for the free recall since the high performers were lost in the analyses due to their lack of incorrect items. The results for the focused questions in Experiment 1 favoured the adults but this effect disappeared when the items that overlapped between the free recall and the focused questions were eliminated from the analysis. Experiment 2 showed no significant differences between the age groups. In brief, summarized over both resolution and slope, the indications of age differences for the focused questions for the discrimination aspect are quite meagre. However, the finding by Roebers and Howie (2003), reviewed above, that for *misleading questions* only adults gave higher confidence judgments for correct answers than for incorrect answers, may be a sign of weakness in children's separation ability, as compared with adults'.

## 6 Level of Noise in Children's and Adults' Confidence Judgments

One possible explanation for the difference observed between children and adults in overconfidence is that children in the investigated ages are *generally less skilled* in using the confidence scale. If children are poor at handling the confidence scale, one might expect the error component in the children's confidence judgments to be affected. The size of the standard deviation for individual participants' confidence judgments can be taken to be an indicator of such an error component. Everything else being equal, when the standard deviation for an individual participant's confidence judgments is higher the error component can be assumed to be greater. In this context it is also relevant that Erev et al. (1994) argued convincingly that greater variability in confidence judgments is associated with greater overconfidence. Thus, taken together and in brief, the notion is that the children's greater overconfidence

**Table 8.7** Mean within-subject standard deviations (and *SD*) for the confidence judgments for the three age groups for Experiment 1 and 2 in the Allwood et al. (2008) study

| | 8–9-year-olds | 12–13-year-olds | Adults | |
|---|---|---|---|---|
| | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *p* |
| *Experiment 1* | | | | |
| Free recall | 14.6 (11.9) | 10.8 (11.1) | 13.6 (9.3) | *ns* |
| Focused questions | 22.1 (9.2) | 18.5 (6.3) | 20.0 (3.7) | *ns* |
| Focused questions when excluding overlapping questions | 22.5 (8.7) | 18.5 (6.4) | 19.3 (3.7) | 1 > 2, $p < 0.05$ |
| *Experiment 2* | | | | |
| Free recall | 10.1 (12.3) | 10.4 (11.1) | 14.5 (12.7) | *ns* |
| Focused questions | 21.8 (9.0) | 20.6 (5.8) | 19.4 (4.1) | *ns* |

*Note*: 1 = 8–9-year-olds; 2 = 12–13-year-olds; 3 = Adults

for focussed questions compared with the adults in previous studies might be explained by the fact that children, due to their lesser skill in handling the confidence scale, have larger individual variability in their confidence judgments, compared with adults.

To explore this possibility Allwood, Granhag, et al. (2006a) computed the standard deviations for each child's confidence judgments in the different scale conditions. The average individual standard deviations in each condition were for the Numerical scale = 16.57, for the Picture scale = 14.64, for the Line scale = 15.82, and for the Verbal scale = 15.29. All these values are lower than the average within-subject standard deviation for the confidence judgments (17.69) in the most similar adult comparison group in our previous research (Allwood et al., 2003; Exp. 2, Phase 1). As shown in Table 8.7, similar analysis of the data reported in Allwood et al. (2008) also did not support this idea since there were no significant differences between the adults and any of the two child groups with respect to these standard deviations.

Thus, taken together, the data reported by Allwood, Granhag, et al. (2006a) and by Allwood et al. (2008) do not support the idea that the children's higher overconfidence compared with the adults can be explained by their poorer ability to handle the confidence rating task as such, at least not as indicated by the presence of larger individual standard deviation for their confidence judgments.

## 7 Discussion

This review of research on the realism in children's confidence judgments of their episodic memory performance has shown that the level of realism in confidence judgments is influenced by a number of different factors. However, the results reviewed also suggest that certain aspects of the confidence scale used may not be as important for the outcome in realism as previously speculated. The results

reported by Allwood, Granhag, et al. (2006a) showed that the realism in children's confidence judgments for focused questions demonstrate a fair amount of stability for the four different types of confidence scales tested.

The review also showed that the results of the comparison of the realism in confidence judgments between children and adults are somewhat complex. For example, the outcome of this comparison depends on the type of memory question given and the aspect of realism investigated. Comparison of the results for children from 8 to 13 years of age and adults (students) for *focused questions* showed fairly clearly that children are more overconfident in their confidence judgments of their own episodic memory.

However, the research in Allwood et al. (2008) showed that the same comparison, when made for free recall performance, did not show any difference between children and adults. Here, if anything, there were indications of lesser overconfidence among the younger children (8–9-year-olds). This was demonstrated in two experiments which only varied in the complexity of the episodic event experienced. We had no formal measure of the complexity of the video clips shown but our strong intuitive impression is that the clip used in Experiment 1 was less complex than that shown in Experiment 2.

Furthermore, the aspect of realism considered makes a difference. The results for the discrimination aspect of metacognitive realism in confidence judgments in the research reviewed did not clearly support any differences between the ability of children in middle childhood and adults.

The study by Allwood et al. (2008) did not solve the question of what aspect of the questions asked that was most causally important for the reported metacognitive results. However, two, probably interacting, aspects appear important. First, there was a difference in the used question formats in the degree of report option allowed (Koriat & Goldsmith, 1996). The free recall instruction gave a large degree of report option, that is, control over what information to report (they assigned a larger answer-area). In contrast, the focused questions involved forced report and assigned a smaller answer-area. When the participants can control what to report (i.e., free recall) they may choose information that they are confident is correct (Koriat et al., 2001). As noted above, this assumes that the participants attempt to hold a high threshold for the level of correctly reported items in their memory report. Even the youngest children in Allwood et al. (2008) appear to have had sufficient training to be able to live up to the philosopher Grice's *quality maxim* for communication, which involves the notion "Try to make your contribution one that is true" (see Schwarz, 1996).

However, apart from the usefulness of report option, it can also be argued that the free recall task allowed access to memories that were easily available. In contrast, in the context of the focused questions, the children were forced to answer questions picked by someone else. On average this can be expected to have had the effect that the focused questions asked the children to provide information that was less accessible in memory compared to the free recall task. The poorer correctness level for these questions compared to the one achieved in the free-recall context in combination with the results for the overconfidence measure, suggests that it may be a harder task to provide realistic confidence judgments for items that are less

accessible in memory. This point was well illustrated above in connection with the results for those of the focused questions that had not been spontaneously mentioned in the participants' free recall in Experiment 1 in the study by Allwood et al. (2008), that is, the non-overlapping questions.

Here it is also relevant to note that a well-known effect in calibration research called the *hard-easy effect* shows that harder questions (poorer correctness) tend to be associated with higher levels of over-/underconfidence (see, e.g., Juslin, Winman, & Olsson, 2000). Future research is needed to systematically sort out the impact of these factors by attempting to vary the correctness level, report option and question format independently of one another.

Finally, the issues of whether some of the scales used in metacognitive confidence research are inappropriate because they are too complex for younger age groups and whether the confidence judgment task as such is too complex for younger children will be discussed. Clearly, there is obviously an age at which a confidence judgment task of one's own memory report is too demanding. However, a number of arguments suggest that these may not be the most pertinent reasons why children of 8–10 years of age have often been found to show worse metacognitive realism than adults. Although children at this age may not understand all aspects of the probability concept, the same can be said of most adults.

Here it is also of interest to note that Schlottmann and Anderson (1994) reported that even 5-year-olds understood the probability concept better than was expected from previous research. The results reported by Allwood et al. (2008) that children of 8–9 years of age showed excellent performance on the overconfidence measure for open free recall, suggest that the confidence judgment task as such may not be too difficult for this age group. Finally, as discussed above, the within-subject standard deviations for the confidence judgments (an indicator of noise in the confidence judgments) did not differ between the four confidence scales investigated by Allwood, Granhag, et al. (2006a), nor between children and adults (Allwood et al., 2003, Exp. 2 phase 1; Allwood, Granhag, et al., 2006a; Allwood et al., 2008).

## 8 Appendix 8.1. Some Common Metacognitive Measures in the Calibration Research Tradition Referred to in the Text

*Calibration* measures the relation between the level of the confidence ratings and the correctness of the memory report. The following formula is used:

$$\text{Calibration} = 1/n \sum_{t=1}^{T} n_t (r_{tm} - c_t)^2$$

Here $n$ is the total number of questions answered. $T$ is the number of confidence classes used; for example if the confidence scale runs from 50% ("guessing") to 100% ("absolutely sure"), the following six confidence classes (T = 6) may be used: 50–59, 60–69, 70–79, 80–89, 90–99, 100). $c_t$ is the percent correct answers of all

items in the confidence class $r_t$ (e.g., 50–59). $n_t$ is the number of times the confidence class $r_t$ was used and $r_{tm}$ is the mean of the confidence ratings in confidence class $r_t$.

*Over-/underconfidence* is computed like calibration, except that the differences are not squared. Over-/underconfidence shows if an individual is overconfident (positive value) or underconfident (negative value). Calibration and over-/underconfidence are perfect when their values are zero.

*Resolution* reflects the ability of the person to discriminate between two sets of answers, one correct and one incorrect. The formula is:

$$\text{Resolution} = 1/n\sum_{t=1}^{T} n_t(c_t - c)^2$$

Here, $c$ is the percent of all items for which the correct answer was provided. A higher value on this measure reflects better resolution than a lower. These measures are better described in Lichtenstein, Fischoff, and Phillips (1982) and in Yates (1994).

*Slope* reflects the ability to separate correct from incorrect answers by means of one's confidence judgments. The formula is:

$$\text{Slope} = (\text{Mean confidence for the correct answers}) - (\text{Mean confidence for the incorrect answers})$$

# 9 Appendix 8.2. The Four Confidence Scales Used in the Study (Translated into English)

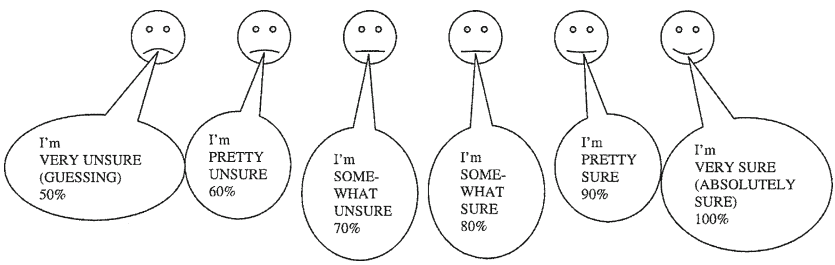
## *9.1 The Numeric Scale*

What was the girl's hair color? (A) Red, (B) Black

I'm _____% sure that I answered the question correct.

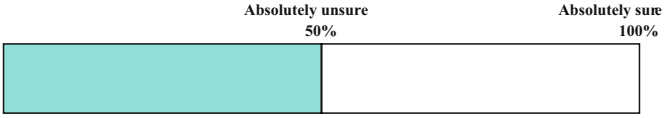
## *9.2 The Picture Scale*

What was the girl's hair color? (A) Red, (B) Black



## *9.3 The Line Scale*

What was the girl's hair color? (A) Red, (B) Black




## *9.4 The Verbal Scale*

What was the girl's hair color? (A) Red, (B) Black

___ 50% Absolutely unsure (Correct 50 times of 100)
___ 60% Pretty unsure (Correct 60 times of 100)
___ 70% Somewhat unsure (Correct 70 times of 100)
___ 80% Somewhat sure (Correct 80 times of 100)
___ 90% Pretty sure (Correct 90 times of 100)
___ 100% Absolutely sure (Correct 100 times of 100)

## References

Allwood, C. M., Ask, K., & Granhag, P. A. (2005). The cognitive interview: Effects on the realism in witnesses' confidence in their free recall. *Psychology, Crime & Law, 11*(2), 183–198.

Allwood, C. M., & Granhag, P. A. (1999). Feelings of confidence and the realism of confidence judgments in everyday life. In P. Juslin & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 123–146). Hillsdale, NJ: Erlbaum.

Allwood, C. M., Granhag, P. A., & Johansson, M. (2003). Increased realism in eyewitness confidence judgments: The effect of dyadic collaboration. *Applied Cognitive Psychology, 17,* 545–561.

Allwood, C. M., Granhag, P. A., & Jonsson A. C. (2006). Child witnesses' metamemory realism. *Scandinavian Journal of Psychology. 47*(6), 461–470.

Allwood, C. M., & Johansson, M. (2004). Actor-observer differences in realism in confidence and frequency judgments. *Acta Psychologica, 117*, 251–274.

Allwood, C. M., Innes-Ker, Å., Homgren, J., & Fredin, G. (2008). Accuracy and realism in confidence in children's and adults' event recall in response to free recall and focused questions. *Psychology, Crime & Law, 14,* 529–547

Allwood, C. M, Jonsson, A. C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' meta-memory accuracy. *Applied Cognitive Psychology, 19,* 331- 344.

Allwood, C. M., Knutsson, J., & Granhag, P. A. (2006). Eyewitness under influence: How feedback affect the realism in confidence. *Psychology, Crime & Law,12,* 25–38.

Christianson, S.-Å., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition & Emotion, 5*(2), 81–108.

Dirkzwager, A. (1996). Testing with personal probabilities: 11-year-olds can correctly estimate their personal probabilities. *Educational and Psychological Measurement, 56*, 957–971.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519–527.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Granhag, P. A. (1997). Realism in eyewitness confidence as a function of type of event witnessed and repeated recall. *Journal of Applied Psychology, 82,* 599–613.

Granhag, P. A., Strömwall, L., & Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitnesses' confidence. *Applied Cognitive Psychology, 14,* 397–420.

Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. *Psychological Review, 104,* 194–202.

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology, 18*(6), 711–726.

Jonsson, A.-C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences, 34,* 559–574.

Jost, J. T., Kruglanski, A. W., & Nelson, T. O. (1998). Social metacognition: An expansionist review. *Personality and Social Psychology Review, 2,* 137–154.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57,* 226–246.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107,* 384–396.

Klayman, J., Soll, J. B., Gonzáles-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what and whom you ask. *Organizational Behavior and Human Decision Processes, 79,* 216–247.

Koriat, A. (2007). Metacognition and consciousness. In P.D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge, UK: Cambridge University Press.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517.

Koriat, A., Goldsmith, M., Schneider, W., & Nakash-Dura, M. (2001). The credibility of children's testimony: Can children control the accuracy of their memory reports? *Journal of Experimental Child Psychology, 79,* 405–437.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Leippe, M.R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. *Law and Human Behavior, 4,* 261–274.

Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. Lindsay, D. F. Ross, J. Don Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology, Volume 2, Memory for people* (pp. 377–425). Mahwah, NJ: Erlbaum.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 306–334). New York: Cambridge University Press.

Marsh, E. J. (2007). Retelling is not the same as recalling: Implications for memory. *Current Directions in Psychological Science, 16*(1), 16–20.

Newman, R. S., & Wick, P. L. (1987). Effect of age, skill, and performance feedback on children's judgments of confidence. *Journal of Educational Psychology, 79*, 115–119.

Nilsson, S.-E. (1998). *Children's decision making under conditions of certainty and risk.* Unpublished doctoral dissertation, Göteborg University, Sweden.

Roebers, C. M. (2002). Confidence judgments in children's and adults' event recall and suggestibility. *Developmental Psychology, 38,* 1052–1067.

Roebers, C. M. (2007). Favourable and unfavourable conditions for children's confidence judgments. *British Journal of Developmental Psychology, 25,* 109–134.

Roebers, C. M., & Howie, P. (2003). Confidence judgments in event recall: Developmental progression in the impact of question format. *Journal of Experimental Child Psychology, 85,* 352–371.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory. *Perspectives on Psychological Science, 1*(3), 181–210.

Sarwar, F., Allwood, C. M., & Innes-Ker, Å. (2010a). *Effects of communication with a non-witness on eyewitnesses' recall correctness and metacognitive realism.* Manuscript submitted for publication.

Sarwar, F., Allwood, C. M., & Innes-Ker, Å. (2010b). Effects of repeated recall and discussion on eyewitness' accuracy and meta-memory realism for different types of forensic information. Unpublished manuscript. Lund University and University of Gothenburg, Sweden.

Schlottmann, A., & Anderson, N. H. (1994). Children's judgments of expected value. *Developmental Psychology, 30,* 56–66.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and the logic of conversation.* Hillsdale, NJ: Erlbaum.

Teigen, K. H. (2001). When equal chances = Good chances: Verbal probabilities and the equiprobability effect. *Organizational Behavior and Human Decision Processes, 85,* 77–108.

Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition, 31,* 1060–1071.

Weinert, F. E., & Kluwe, R. H. (Eds.). (1997). *Metacognition, motivation and understanding.* Hillsdale, NJ: Erlbaum.

Yates, J. F. (1994). Subjective probability accuracy analysis. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 381–410). New York: Wiley.