The Effect of Feedback on Children's Metacognitive Judgments:

A Heuristic Account

Marie Geurten & Thierry Meulemans

Department of Psychology, Psychology and Neuroscience of Cognition Unit, University of Liège, Belgium

E-mail: Marie Geurten, mgeurten@ulg.ac.be
Thierry Meulemans, thierry.meulemans@ulg.ac.be

Correspondence concerning this article should be addressed to Marie Geurten, University of Liège, University of Liège B33 Trifacultaire - Quartier Agora, Place des Orateurs 1, 4000 Liège – Belgium; E-mail: mgeurten@ulg.ac.be; Phone number: +32 4 366 59 43

Acknowledgment: This research was supported by a grant from the Marie Curie Cofund Program to Marie Geurten. We have no conflict of interest to declare.

Abstract

In three experiments, we investigated whether the feedback effect on the accuracy of children's metacognitive judgments results from an improvement in monitoring processes or the use of the Anchoring-and-Adjustment heuristic. Experiment 1 revealed that adding feedback increased the accuracy of young children's (aged 4, 6, and 8 years) memory prediction. In Experiment 2, the influence of an external anchor on children's metacognitive judgment was established. Finally, in Experiment 3, two memory tasks that differed in terms of difficulty were administered. Participants were randomly assigned to an anchoring (high/low/no anchor) and a feedback (feedback/no feedback) condition. Results demonstrated that children in the feedback condition adjusted their predictions toward the feedback, regardless of the task's difficulty. These findings are consistent with the hypothesis that external information provided by feedback is used as an anchor for judgment. This interpretation is strengthened by the correlation found between the two scores computed to assess participants' susceptibility to anchoring and feedback effects, which indicates that children who are more sensitive to the anchoring effect are also more sensitive to the feedback effect.

Keywords: Metacognition, Judgment, Feedback, Anchoring, Children

The Effect of Feedback on Children's Metacognitive Judgments: A Heuristic Account

Decades of intensive studies on metacognition have established that how accurately people assess their memory determines how efficiently they regulate their learning (Dunlosky & Hertzog, 1998; Kornell & Bjork, 2007; Kornell & Metcalfe, 2006; T. O. Nelson & Narens, 1994; Son, 2010; Thiede, 1996). Usually, the ability to assess internal cognitive processes (i.e., metacognitive monitoring) is appraised by asking participants to judge the quality of their performance before (prospective judgment) or after (retrospective judgment) carrying out a cognitive task. These metacognitive judgments can be made on an item-by-item basis (local level) or for the task as a whole (global level). The accuracy of judgments can be assessed with calibration scores (Van Overschelde & Nelson, 2006). This type of score shows the degree to which the judgment differs from the actual level of performance.

Children's metacognitive judgments are far from accurate. Many developmental studies have demonstrated that children regularly overestimate their memory performance when making metacognitive judgments (e.g., Kvavilashvili & Ford, 2014; Lipko, Dunlosky, Lipowski, & Merriman, 2012). Interestingly, this effect seems to be particularly pronounced when the accuracy of global prospective judgments is examined (Lipko, Dunlosky, & Merriman, 2009; Shin, Bjorklund, & Beck, 2007; Yussen & Levy, 1975). For instance, Shin et al. (2007) presented three groups of children (aged 6, 7, and 9 years old) with a set of pictures, then asked them to estimate how many items they thought they would be able to remember on a subsequent test (global prospective judgment). The results revealed that more than 75% of the 6- and 7-year-old children overestimated the number of pictures they would recall. Moreover, at 9 years of age, almost 50% of children still overestimated their memory performance. Given the amount of new learning that children have to master each day and the well-established influence of accurate monitoring processes on cognitive and academic performance (Everson & Tobias, 1998; Pintrich & de Groot, 1990; Roebers, Schmid, & Roderer, 2009), research on metacognition has long focused on identifying methods and situations that could improve children's monitoring abilities (e.g., Koriat, Goldsmith, Schneider, & Nakash-Dura, 2001).

A procedure that has been assumed to increase the accuracy of both adults' and children's monitoring processes involves providing external feedback to draw participants' attention to the fact that their performance was evaluated as poor (Butler, Karpicke, & Roediger, 2008; Efklides & Dina, 2004; Kornell & Rhodes, 2013; Metcalfe & Finn, 2011, 2012; Miller & Geraci, 2011; Roll, Aleven, McLaren, & Koedinger, 2011). Nevertheless, although the positive influence of external feedback on adults' (Metcalfe & Finn, 2011) and older children's (Metcalfe & Finn, 2012) metacognitive judgments seems to be well established, to our knowledge, this effect has never been studied in participants under the age of 8 years. Yet, according to the literature, young children tend to have the highest level of metacognitive overestimation; thus, they are certainly in the greatest need of corrective feedback to improve the accuracy of their predictions (Lipko et al., 2009; Shin et al., 2007; Yussen & Levy, 1975).

For these reasons, the first aim of the present study is to investigate whether young children are able to make more accurate global prospective judgments after receiving concrete feedback about their previous metacognitive predictions (Experiment 1). According to Koriat (2007), participants may not always rely on effortful monitoring processes to make their judgments; they may also base their decisions on various metacognitive heuristics (i.e., automatic inference rules that are used to guide memory decisions on the basis of a variety of cues. These rule are supposed to operate at the fringe of consciousness; Reber, Schwarz, & Winkielman, 2004). For instance, Koriat, Ackerman, Lockl, and Schneider (2009a, 2009b) have shown that 9-year-old children's metacognitive judgments are based on the automatic inference that easily learned items are more likely to be remembered (i.e., memorizing-effort heuristic), suggesting that metacognitive heuristics can be used by children to evaluate their own learning. Although Koriat and colleagues did not detect any use of this heuristic in younger children, studies have recently indicated that children could already rely on this automatic inference rule as early as 4 (Geurten, Willems, & Meulemans, 2015). We therefore postulated that the Anchoring-and-Adjustment heuristic could potentially account for the effect of feedback on metacognitive judgments.

The anchoring effect is defined as the adjustment – higher or lower – of the estimation of an unknown quantity, based upon previously presented external information, namely the anchor (e.g., Epley & Gilovich, 2006; McElroy & Dowd, 2007; Simmons, LeBoeuf, & Nelson, 2010; Tversky & Kahneman, 1974). Tversky and Kahneman (1974) were the first to demonstrate the influence of the anchoring effect on people's judgments. In their classic study, they found that adults generally estimated that a greater percentage of African countries belong to the United Nations after being exposed to an external anchor of 65% than after being exposed to an external anchor of 10%. Over time, the Anchoring-and-Adjustment heuristic has been claimed to be involved in a wide range of numerical judgments (Englich, Mussweiler, & Strack, 2006; Northcraft & Neale, 1987; Switzer & Sniezek, 1991; Wright & Anderson, 1989), including the appraisal of future memory performance (Scheck & Nelson, 2005). To date, however, the ability to rely on the heuristic to guide judgments has not been studied in children under the age of 9 years (Smith, 1999). A second experiment was therefore conducted to explore whether young children (aged 4, 6, and 8 years old) adjust their metacognitive judgments upward or downward after being randomly presented with a high or low anchor (Experiment 2). Specifically, in our second experiment, children were asked whether they thought they would remember more or fewer than 12 words (high anchor) or more or fewer than 2 words (low anchor). If children make their memory decisions on the basis of the Anchoring-and-Adjustment heuristic, we expect them to predict that they will be able to recall more words when a high anchor is provided than when no anchor is given. Conversely, we expect them to predict that they will be able to recall fewer words when a low anchor is provided than when no anchor is given.

Finally, we examine whether the influence of feedback on metacognitive judgment is due to an anchoring effect. To this end, a third experiment was carried out in which two memory tasks were administered to children. Half of the participants received feedback about the accuracy of their judgment after the first task. This time, however, we adopted a procedure that was anticipated to lead to different results depending on whether feedback induced children to engage in effortful metacognitive processes or whether it was used as a cue for automatic inference. Specifically, we presented children with two tasks that varied in terms of their level of difficulty (easy or hard). The rationale for this choice was that participants' performance was supposed to vary as a function of task difficulty. Consequently, feedback on the number of items recalled during the first task was not an appropriate cue to predict performance on the second task. If the feedback effect results from the implementation of the

Anchoring-and-Adjustment heuristic, then participants are expected to adjust their global prediction based on the feedback they are given, regardless of differences in task difficulty. Indeed, if an easy task is administered first, children will perform well and, thus, will receive positive feedback (high anchor). According to the anchoring hypothesis, we expect participants to predict that they will perform better on the second (hard) task than children who receive no feedback. Similarly, if a hard task is given first, children will perform poorly and, thus, will receive negative feedback (low anchor). If feedback is used as an anchor, we anticipate participants to predict that they will remember fewer items for the second (easy) task than children in the no feedback condition. Conversely, if the effect of feedback is to draw children's attention to the fact that they had evaluated their memory poorly, participants in the feedback condition are expected to take the difficulty of the task into account when making a subsequent judgment, resulting in more accurate judgments for children in the feedback condition than for children in the no feedback condition.

To sum up, the main goals of this study are to (a) determine whether getting feedback about the accuracy of the metacognitive judgment allows young children to better predict their future memory performance; (b) examine whether preschool and early school-aged children rely on the Anchoring-and-Adjustment heuristic to guide their memory decisions; and (c) explore whether the processes underlying the effect of feedback on children's metacognitive judgments results from an anchoring effect (where feedback serves as an external anchor).

Experiment 1

The purpose of Experiment 1 was to investigate whether 4-, 6-, and 8-year-old children are able to use feedback to increase the accuracy of their memory judgments. To do so, children in three age groups were divided into two experimental conditions (feedback or no feedback). After studying a list of associated words, participants were instructed to predict their future memory performance, and then they were asked to recall as many items as possible. Next, half of the participants were given concrete feedback about the accuracy of their global prediction. Once the feedback was provided, all children were presented with another set of associated word pairs and the procedure was repeated. We expected children in the feedback condition to make more accurate prospective judgments than children in the no feedback condition, regardless of their age. This pattern would demonstrate young children's susceptibility to feedback effects. From a developmental point of view, the changes in this feedback effect across age groups were also examined.

Method

Participants. Participants were 48 typically developing children aged 4 (*n* = 16; Mean = 52.99 months, SD = 5.41), 6 (*n* = 16; Mean = 76.20 months, SD = 5.37), and 8 (*n* = 16; Mean = 101.40 months, SD = 3.45) years old. The proportion of girls and boys was strictly equivalent in each group. The mean of both parents' years of education was used to appraise socioeconomic status (Mean = 13.78, SD = 2.03), and standard scores on the Matrix subtest of the Wechsler Intelligence Scale for Children (WISC-IV; Wechsler, 2005) and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III; Wechsler, 2004) were used to evaluate nonverbal intelligence (Mean = 10.31, SD = 2.66). No group difference was found between the three age groups for socioeconomic status and nonverbal intelligence, *Fs*<1, *ps*>.52. The sample was recruited from French-speaking kindergartens and elementary schools in the province of Liège, Belgium.

Materials. Stimuli consisted of 30 pairs of words (cue-target). All cues and targets were French nouns between three and nine letters in length (Mean = 5.47) and with one to three syllables (Mean = 1.73). Each word selected was expected to be included in the vocabulary of 4-year-old children. The pairs were matched so that no easy semantic or phonological association could be made between the two items in a given pair (e.g., *Cloud-Key*). Half of the 30 pairs were randomly assigned to one of the two experimental lists. The two sets of items were equal in terms of word frequency (10.7 and 9.1 occurrences per million words; Radeau, Mousty, & Content, 1990), length (a mean of 5.33 and 5.60 letters per word), and number of syllables (a mean of 1.66 and 1.8 syllables). A pretest carried out on 15 participants (aged 4, 6, and 8 years old) indicated that the two lists were equal in terms of the number of words that children were able to recall after one trial in a cued-recall task.

Procedure and design. The study was approved by the local ethics committee. Written consent was obtained from the parents before the study began. Children were tested individually in a quiet room in their school. Each child underwent a 45-minute session during which the two verbal memory tasks were administered. The order of these tasks was counterbalanced between subjects. All participants were informed that they would have to study 15 cue-target pairs, make a global prospective judgment and take a cued-recall test; this procedure would take place twice. The study and recall phases of each task were separated by a 10-minute delay that was filled with two nonverbal cognitive tasks. One of these tasks was the Matrix test. Figure 1 illustrates the procedure of Experiments 1 to 3.

A 3 (Age Group: 4-, 6-, or 8-year-olds) x 2 (Condition: feedback or no feedback) between-subjects design was used in Experiment 1. There were 8 children per cell.

Study phase. A list of 15 word pairs was presented twice in random order to each child. Participants were asked to try to remember as many pairs as possible so they would be able to recall as many targets as possible when the associated cues were presented later. More specifically, participants were asked to stare at a fixation cross (+) in the center of the screen while the two items were named by a female voice (over the computer speakers) at a rate of one every 4 s. Because the children in the two younger groups were not able to read, we chose not to accompany the auditory presentation of the words with a visual one (i.e., written words) to avoid favoring the older children. Once every pair had been given, the items were presented for re-study. This procedure was chosen because a pretest indicated that most 4-year-old children were not able to recall more than one target after a single presentation of the pairs.

Judgment phase. The study phase was immediately followed by a metacognitive judgment solicitation. Specifically, children were asked to estimate how many of the 15 targets they thought they would remember in response to the presentation of the cue in an upcoming test (i.e., global prospective judgment).

Recall phase. A 10-minute delay followed the judgment phase. Once this time had elapsed, there was a final cued-recall test. The task was to recall the second word of each pair in response to the presentation of the first one. There was no time limit to complete the task. At the end of the test, half of the participants were given concrete feedback regarding their prospective judgment (feedback condition). Specifically, children were first reminded of their metacognitive prediction regarding their memory performance (e.g., “A bit earlier, you said you would recall 9 items out of the 15 you studied”). Then they were told the number of pairs they were actually able to remember (e.g., “In fact, you recalled 6 words”). No metacognitive feedback was provided to the other half of the sample (no feedback condition). The procedure described above was repeated for the second memory test (see Figure 1).

< Figure 1 about here >

Measures. For each of the two memory tasks, the recorded measures were (a) the number of correct responses for the cued-recall test (memory score); (b) the number of items predicted during the judgment phase (judgment score); and (c) the accuracy of children's predictions compared with their actual memory performance (calibration score). The latter score was computed using a formula adapted from the one developed by Hacker, Bol, and Bahbahani (2008). Using this equation, a score of 0 indicates perfect accuracy, a negative score indicates that children underestimated their memory performance, and a positive score indicates that children overestimated their memory performance.

$$Calibration = \left(\frac{prediction - performance}{total\ \ items} \right)$$

Results

Data analyses. A 3 (Age Group: 4-, 6-, or 8-year-olds) × 2 (Condition: feedback or no feedback) between-subjects analysis of variance (ANOVA) was conducted to investigate the effect of metacognitive feedback on children's memory predictions. As the feedback effect could not logically be observed on the memory task that was presented first, statistical analyses were only carried out on the scores for the second memory task (i.e., the task performed after the feedback was given). Moreover, to ensure that the difference observed between the feedback and no feedback conditions for the second memory task was not due to initial differences in participants' cognitive and metacognitive abilities, the memory, judgment, and calibration scores recorded during the first task were included as covariates in the analyses carried out on the memory, judgment, and calibration scores, respectively, recorded for the second task. The significance level was set at .05. Statistical analyses indicated homogeneity of variance between age groups and feedback conditions, and revealed no gender or order effect on any of the dependent variables. Table 1 displays the mean number of correct responses, as well as the number of items predicted and the accuracy of predictions for each experimental condition (feedback/no feedback).

Memory score. We first examined whether the feedback condition affected children's memory performance. The results of an ANCOVA revealed a main effect of age, $F(2,42) = 4.71$, $MSe = 2.00$, $p = .01$, $\eta^2_p = .19$, on the number of targets recalled, indicating that older children had better memory performance than younger children. However, no other effect reached significance, $Fs<1.60$, $ps>.21$.

Judgment score. The effect of feedback on children's global memory prediction was analyzed. No age effect or age x condition interaction was found, $Fs<1$, $ps>.77$. However, the results revealed a main effect of condition, $F(1,42) = 8.75$, $MSe = 5.89$, $p = .005$, $\eta^2_p = .18$, indicating that children's predictions were lower in the feedback (Mean = 5.88) than in the no feedback condition (Mean = 7.95). The effect of the covariate (judgment score for the first task) was also significant, $F(1,42) = 26.96$, $MSe = 5.89$, $p < .001$, $\eta^2_p = .40$.

< Table 1 about here >

Calibration score. Finally, the effect of feedback condition and children's age on the accuracy of participants' memory prediction was examined. The results of the ANCOVA revealed that children in the feedback condition showed greater metacognitive calibration than children in the no feedback condition, $F(1,42) = 10.23$, $MSe = 0.04$, $p = .003$, $\eta^2_p = .20$. An effect of the covariate (calibration score for the first task) was also found, $F(1,42) = 19.40$, $MSe = 0.04$, $p < .001$, $\eta^2_p = .33$. No age effect or age x condition effect was found, *Fs*<1, *ps*>.56. On the whole, these results seem to suggest that all children were able to use the feedback to improve the accuracy of their future metacognitive judgments.

Discussion

The main goal of Experiment 1 was to demonstrate the effect of feedback on young children's metacognitive judgments. Our results revealed that children's predictions were more accurate in the feedback than in the no feedback condition, indicating that getting feedback about the accuracy of their judgments had a positive influence on their subsequent memory predictions. Moreover, in this experiment, most of the participants (95.9%) overestimated their future memory performance. Therefore, children in the feedback condition generally received negative feedback about their predictions. This could have affected their motivation to do the task and could therefore have reduced their memory performance (for a meta-analysis investigating the effect of negative feedback on cognitive performance, see Kluger & DeNisi, 1996). However, although the feedback was shown to affect the metacognitive scores (i.e., prediction and calibration scores) by inducing participants to revise their judgment downward to match their recent memory performance, no effect was observed on objective memory performance.

From a developmental perspective, the results showed that the calibration of global prospective judgments improves with age (Shin et al., 2007). Regarding the feedback effect, however, our findings revealed that all children – whatever their age – were sensitive to the influence of feedback in the same way.

As the sensitivity of young children to feedback was now established, we examined the nature of the processes underlying this feedback effect. According to the literature, automatic decision rules have been demonstrated to be involved in people's judgments (for an overview, see Koriat, 2007). One of these automatic rules, the Anchoring-and-Adjustment heuristic, seems to be frequently used when numerical judgments are required (Tversky & Kahneman, 1974). However, to our knowledge, the involvement of this heuristic in preschoolers' and early school-aged children's memory judgments has never been studied. Experiment 2 was therefore carried out to examine young children's ability to base their prospective judgments on an external anchor. Once the anchoring effect was demonstrated, a third experiment was conducted to test whether the anchoring effect accounted for the influence of feedback on memory predictions.

Experiment 2

In Experiment 2, we sought to determine whether young children are able to use the Anchoring-and-Adjustment heuristic to guide their global memory predictions. For this purpose, 4-, 6-, and 8-year-old children were divided into three experimental conditions depending on the anchor that was provided (high, low, or no anchor). We predicted that children would adjust their prospective judgments in the direction of the anchor they were given, demonstrating their susceptibility to anchoring effects. Furthermore, the developmental trend of children's ability to rely on the heuristic was also explored.

Method

Participants. A total of 45 typically developing unilingual children aged 4 (Mean = 55.07 months, SD = 2.88, n = 15, 8 females), 6 (Mean = 82.59 months, SD = 2.98, n = 15, 8 females), and 8 (Mean = 103.97 months, SD = 3.08, n = 15, 7 females) years old participated in Experiment 2. No group difference was found in terms of parental education level and nonverbal intelligence, $Fs<1.50$, $ps>.23$.

Materials. A list of 15 unrelated pairs of French words (cue-target) was used in Experiment 2. This list was one of the two used in Experiment 1.

Procedure and design. Each participant underwent a 30-minute session. In Experiment 2, only one memory task was administered (see Figure 1). Once again, the study and recall phases were separated by a 10-minute delay that was filled with two nonverbal cognitive tasks, including the Matrix test. The procedure was identical to the one described in Experiment 1 except for the judgment phase. In that phase, participants were divided into three experimental conditions: high, low, or no anchor. In the high anchor condition, children were asked to say whether they thought they would remember more or fewer than 12 targets; this initial information was used as the anchor. After that, participants were asked to estimate the exact number of items they thought they would remember. The same procedure was employed in the low anchor condition, except that children were first asked whether they thought they would be able to recall more or fewer than 2 targets. In the no anchor condition, children were simply instructed to estimate the number of targets they would be able to recall in the upcoming cued-recall test. Thus, a 3 (Age Group: 4-, 6-, or 8-year-olds) x 3 (Condition: high, low, or no anchor) between-subjects design was used in Experiment 2. There were 5 children per cell.

Measures. The key analyses conducted in Experiment 2 focused on (a) the number of correct responses for the cued-recall test (memory score); (b) the number of items predicted (judgment score); and (c) the accuracy of children's predictions compared with their actual memory performance, assessed using the same formula as in Experiment 1 (calibration score).

Results

Data analyses. For each dependent variable, a 3 (Age Group: 4-, 6-, or 8-year-olds) × 3 (Condition: high, low, or no anchor) between-subjects ANOVA was carried out to explore whether young children adjusted their prospective judgments upward or downward because of an anchor, demonstrating their ability to rely on the Anchoring-and-Adjustment heuristic. Preliminary analyses indicated homogeneity of variance between age groups and experimental conditions. No gender effect was revealed on any of the dependent variables. Table 2 displays the mean number of correct responses, as well as the number of items predicted and the calibration of predictions in each experimental condition (high, low, or no anchor).

Memory score. As in Experiment 1, we first investigated the influence of age and experimental condition on participants' memory performance. The results of the two-way ANOVA showed a main effect of age, $F(2,40) = 8.54$, $MSE = 3.58$, $p < .001$, $\eta^2_p = .32$, with older children obtaining higher memory scores than younger participants. However, no effect of anchoring condition and no age x condition interaction was found, $Fs<1.40$, $ps>.26$.

Judgment score. Examination of the effect of age and anchoring condition on children's global prospective judgment revealed a main effect of experimental condition, $F(2,40) = 7.03$, *MSe* = 7.00, $p = .003$, $\eta^2_p = .28$. Linear contrast analyses indicated that children in the low anchor condition predicted that they would remember fewer targets (Mean = 5.27) than children in the high (Mean = 8.93) and no anchor (Mean = 7.47) conditions, $p \leq .032$. However, no statistically significant difference was found between the high anchor and no anchor conditions, $p = .14$. No other main effect or interaction of the two-way ANOVA reached significance, *Fs*<1, *ps*>.60. These findings suggest that the ability to rely on an anchor to make memory judgments is acquired early in childhood.

< Table 2 about here >

Calibration score. The results of the 3 (Age Group) x 3 (Condition) ANOVA revealed no interaction effect, $F(2,40) = 0.73$, *MSe* = 0.04. However, a main effect of age was found, $F(2,40) = 5.17$, *MSe* = 0.04, $p = .01$, $\eta^2_p = .22$, indicating that older children predicted their memory performance more accurately than younger children. A statistical trend toward an effect of the experiment condition was also found, $F(2,40) = 2.75$, *MSe* = 0.04, $p = .08$, $\eta^2_p = .13$. Linear contrast analyses revealed that children made a more calibrated judgment when a low anchor was provided (Mean = 0.11) than when a high anchor was provided (Mean = 0.28), $p = .02$. However, no difference was found in the no anchor condition, all *ps*>.20. This pattern of results is coherent with the finding that children generally overestimate their performance when making prospective judgments (Shin et al., 2007). Because providing a low anchor leads participants to revise their predictions downward, it results in more accurate judgments than when a high anchor is given.

Discussion

The anchoring effect is frequently used to account for a wide variety of numerical judgments. However, to our knowledge, young children's ability to use an external anchor to guide their decisions has never been studied. This experiment demonstrated that children's predictions were higher in the high than in the low anchor condition, suggesting that children in all age groups adjusted their prospective judgment depending on the random anchor they were given. As children generally overestimate their memory performance when making a prediction (e.g., Shin et al., 2007), we are not surprised to observe a larger adjustment of the judgment in the low anchor condition than in the high anchor condition. Nevertheless, the latter finding indicates that the ability to rely on the Anchoring-and-Adjustment heuristic develops early. As young children's ability to make memory predictions on the basis of an external anchor was now established, an additional experiment was carried out to examine the possible involvement of the Anchoring-and-Adjustment heuristic in the feedback effect.

Experiment 3

Experiment 1 confirmed that receiving feedback about the accuracy of a previous metacognitive judgment helps participants make a better memory prediction on a subsequent test. The primary focus of the present study was to determine what mechanisms underlie this feedback effect. Specifically, we postulate that the Anchoring-and-Adjustment heuristic could also account for the influence of feedback on memory prediction. In this context, the results of Experiment 2 demonstrating that young children are able to rely on an external anchor to guide their metacognitive judgment provide preliminary, but not sufficient, evidence for the latter hypothesis.

For these reasons, the primary aim of Experiment 3 was to determine whether the feedback effect can serve as an external anchor for children's global prospective judgments (Scheck & Nelson, 2005). To do so, children were presented with two memory tasks that differed in terms of difficulty. Participants were randomly assigned to an anchoring (high, low, or no anchor) and a feedback (feedback or no feedback) condition to obtain a balanced experimental design. Assuming the anchoring hypothesis was correct, we anticipated that children would demonstrate less calibrated prospective judgment in the feedback than in the no feedback condition because they would adjust their global predictions in the direction of the feedback provided, regardless of the task's difficulty (e.g., receiving positive feedback or a high anchor following an easy task would lead children to wrongly make a high prospective judgment for the hard subsequent task). Moreover, the relationship between participants' vulnerability to anchoring effects and their vulnerability to feedback effects was also explored. The aim of the latter analysis was to show that children who adjusted their prediction toward the anchor in the first task were also more likely to adjust their prediction toward the feedback in the second task. Such a pattern would provide further evidence in favor of the anchoring hypothesis because it would indicate that the same processes influence children's judgment when they are confronted with an anchor or with feedback.

Method

Participants. Participants were 108 typically developing children aged 4 (*n* = 36; Mean = 53.88 months, SD = 3.87), 6 (*n* = 36; Mean = 79.44 months, SD = 3.91), and 8 (*n* = 36; Mean = 102.39 months, SD = 3.58) years old. Fifty percent of the subjects were girls. The groups were roughly equivalent in terms of socioeconomic status (parental education level) and nonverbal intelligence (Matrix score), *Fs*<1, *ps*>.59.

Materials. Two sets of 15 word pairs included in the vocabulary of 4-year-old children served as critical items. All cues and targets were French nouns between three and nine letters long (Mean = 5.45) and with one to three syllables (Mean = 1.65). Half of the pairs were created so that no easy semantic or phonological association could be found between the two items in a given pair (e.g., *Cloud-Key*). The remaining items were matched so that a quick semantic association could be made between the cue and the target (e.g., *Day-Night*). The procedure developed by D. Nelson, McEvoy, and Schreiber (2004) was used to determine the strength of the association between the two items in each pair. Ten related pairs and five unrelated pairs were included in the "easy" list. Ten unrelated pairs and five related pairs were included in the "hard" list. The two sets of items were equal in terms of word frequency (11.3 and 14.3 occurrences per million words), length (a mean of 5.48 and 5.31 letters per word), and number of syllables (a mean of 1.65 and 1.72 syllables). The two lists of word pairs are presented in the Appendix. The results of a pretest indicated that children produced significantly more correct responses after studying the easy list (10 related pairs) than after studying the hard list (5 related pairs), demonstrating that the former really was easier than the latter.

Procedure and design. Each child participated in a 45-minute session. Participants were instructed that they would be asked to study 15 cue-target pairs, give a global prospective judgment and perform a cued-recall test; they would go through this procedure twice (see Figure 1). The procedure for the study and recall phases of the two memory tasks was identical to the one described in Experiment 1 (feedback procedure). The judgment phase for the first set of pairs was the same as in Experiment 2 (anchoring procedure). However, the judgment phase for the second set of pairs was the same as in Experiment 1. We did not provide an anchor before the second prediction phase in order to examine the effect of feedback without interference. Specific instructions for each experimental condition are provided as supplemental material. The order of the two memory tasks (easy vs. hard) was counterbalanced between subjects.

Thus, a 3 (Age Group: 4-, 6-, or 8-year-olds) x 3 (Anchoring: high, low, or no anchor) x 2 (Feedback: feedback or no feedback) x 2 (Task Difficulty: easy or hard) mixed-subjects design was used in Experiment 3. Task difficulty was the only within-subject factor. Specifically, an easy task and a hard task were administered to all participants, but the order of these tasks was counterbalanced between subjects. Each child was randomly assigned to one of the anchoring and feedback conditions so as to obtain a balanced design. There were 6 children per cell.

Measures. For each memory test, the main dependent variables included in our analyses were (a) the number of correct responses (memory score); (b) the number of items predicted (judgment score); and (c) the accuracy of children's predictions as a function of their actual memory performance (calibration score). In addition, a score for susceptibility to anchoring effects was computed for the participants who received an anchor before making their prospective judgment. This score was computed using the measures recorded during the first memory task (i.e., the task for which an anchor was given). In the high anchor condition, the mean number of words predicted by children to whom no anchor was given (Mean = 4.71) was subtracted from the number of items predicted by each child. In the low anchor condition, the number of items predicted by each child was subtracted from the mean number of words predicted by the children included in the no anchor condition. In this context, a positive score indicates that the children adjusted their prediction toward the anchor whether it was high or low. Finally, a score for susceptibility to feedback effects was calculated to determine the extent to which children adjusted their predictions after receiving feedback about the accuracy of their previous judgments. First, a discrepancy score between the number of items predicted for the first and second memory tasks was computed for each participant. Then the mean discrepancy score for children included in the no feedback condition (Mean = 2.25) was subtracted from the discrepancy score of each child in the feedback condition. A positive score indicated that participants adjusted their predictions more radically after receiving feedback than they did if no feedback was given.

Results

Data analyses. The primary aim of this third experiment was to determine whether feedback is used as a cue for the Anchoring-and-Adjustment heuristic (anchoring hypothesis). To examine the occurrence of an anchoring effect, a 3 (Age Group: 4-, 6-, or 8-year-olds) × 3 (Anchoring: high, low, or no anchor) x 2 (Task Difficulty: easy or hard) mixed-subjects ANOVA was conducted on the number of correct responses, the number of items predicted, and the calibration of predictions for the first memory task. Task difficulty was the only within-subject factor. Because no anchor was provided before the second memory test, the anchoring effect was only investigated for the data recorded during the first memory test.

Next, the feedback effect was investigated. For this purpose, a 3 (Age Group: 4-, 6-, or 8-year-olds) × 2 (Feedback: feedback or no feedback) x 2 (Task Difficulty: easy or hard) ANCOVA was carried out on the number of correct responses, the number of items predicted, and the calibration of predictions for the second memory task. As in Experiment 1, the influence of memory, judgment, and calibration scores recorded during the first memory task was taken into account in the following analyses. A significant feedback x task difficulty interaction would indicate that children based their prospective judgments on the feedback, but did not adjust their predictions depending on task difficulty. This pattern would thus provide evidence in favor of the anchoring hypothesis. Because the feedback could not affect performance on the first memory task, statistical analyses were not carried out on the data recorded during that test.

Finally, to explore the relationship between children's vulnerability to anchoring effects and their vulnerability to feedback effects and, thus, to determine whether these effects share common processes, a Pearson's correlation analysis was computed on the two susceptibility scores. Moreover, we also examined whether susceptibility to anchoring and feedback effects varied with the children's age. To this end, Pearson's correlation analyses were carried out between the two susceptibility scores and participants' chronological age (in months). In the following sections, the significance level was set at .05.

Preliminary analyses. Preliminary analyses indicated homogeneity of variance, and revealed no gender or order effect on any of the dependent variables. Moreover, to ensure that the anchoring manipulation introduced during the judgment phase of the first memory task had no delayed effect on the scores recorded during the second memory test, a 3 (Anchoring) x 2 (Feedback) ANOVA was carried out on the memory, judgment, and calibration scores of the second memory task. No main effect or interaction was found on any of the dependent measures, *Fs*<1, *ps*>.45, indicating that providing an anchor during the first phase had no late effect on children's performance.

Anchoring effect.

Memory score. The results of the three-way ANOVAs conducted on the number of correct responses revealed a significant main effect of age, *F*(2,102) = 33.40, *MSe* = 2.59, $p < .001$, $\eta^2_p = .43$, indicating that older children had better memory performance than younger children. Furthermore, the number of correct responses also appeared to be higher for the easy task than for the hard task, *F*(1,102) = 84.94, *MSe* = 2.59, $p < .001$, $\eta^2_p = .49$. No other main effect or interaction reached significance, *Fs*<2, *ps*>.13.

Judgment score. Replicating the results of Experiment 2, the ANOVA carried out on the judgment score revealed a main effect of anchoring condition, *F*(2,102) = 27.49, *MSe* = 6.05, $p < .001$, $\eta^2_p = .37$. Linear contrast analyses indicated that participants in the high anchor condition predicted that they would remember more targets (Mean = 7.83) than participants in the no anchor (mean = 5.05) condition, $p < .001$, and that participants in the low anchor (Mean = 3.64) predicted that they would remember fewer words than participants in the no anchor condition, $p = .016$. No other effect reached significance, *Fs*<2.56, *ps*>.08.

Calibration score. As can be seen in Table 3, the results of the ANOVA conducted on calibration scores revealed a main effect of age, *F*(2,102) = 4.89, *MSe* = 0.04, $p = .01$, $\eta^2_p = .10$, indicating that older participants made more calibrated judgments than younger participants. A main effect of task difficulty was also found, *F*(1,102) = 22.17, *MSe* = 0.04, $p < .001$, $\eta^2_p = .20$. More specifically, children’s predictions for the easy task were more calibrated than their predictions for the hard task, which is consistent with the fact that memory performance was higher for the easy task than for the hard task while no differences were found between these two tasks for the judgment score. Moreover, a main effect of anchoring condition was also found, *F*(2,102) = 22.54, *MSe* = 0.04, $p < .001$, $\eta^2_p = .33$. Linear contrast analyses revealed a significant difference between each of the three conditions (high, low, or no anchor), all *ps* < .02, with better prediction calibration in the low anchor condition than in the no anchor condition and a better calibration in the no anchor condition than in the high anchor condition.

< Table 3 about here >

Feedback effect.

Memory score. The three-way ANCOVA demonstrated that participants recalled more targets for the easy than for the hard task, *F*(1,102) = 66.64, *MSe* = 2.49, $p < .001$, $\eta^2_p = .41$. A main effect of age was also found, *F*(2,102) = 3.64, *MSe* = 2.49, $p = .03$, $\eta^2_p = .07$: older children recalled more words than younger children. The effect of the covariate (memory score for the first task) was also significant, *F*(2,102) = 26.75, *MSe* = 2.49, $p < .001$, $\eta^2_p = .22$. No other main effect or interaction was found, all *Fs*<1.03, *ps*>.31.

Judgment score. The influence of age group, feedback condition, and task difficulty on children's global prospective judgment was examined. The results of the statistical analyses showed a main effect of age, *F*(2,95) = 10.29, *MSe* = 5.59, $p < .001$, $\eta^2_p = .18$: younger children made lower predictions than older children. The two older groups did not differ in terms of the number of items predicted, $p = .37$. As Figure 2 shows, a feedback x task difficulty interaction was also found, *F*(2,95) = 19.56, *MSe* = 5.59, $p < .001$, $\eta^2_p = .17$. In the feedback condition, higher predictions were made for the hard task (Mean = 6.74) than for the easy task (Mean = 4.81), *F*(1,95) = 8.15, *MSe* = 5.59, $p = .005$. Conversely, in the no feedback condition, lower predictions were made for the hard task (Mean = 4.85) than for the easy task (Mean = 6.70), *F*(1,95) = 7.53, *MSe* = 5.59, $p = .006$. Moreover, children also made higher predictions for the hard task in the feedback (6.74) than in the no feedback condition (4.85), *F*(1,95) = 7.84, *MSe* = 5.59, $p = .006$, while they made lower predictions for the easy task in the feedback (4.81) than in the no feedback condition (6.70), *F*(1,95) = 7.83, *MSe* = 5.59, $p = .006$. These findings suggest that children in the feedback condition adjusted their predictions toward the feedback they were given, regardless of the difficulty of the task. On the contrary, children in the no feedback condition seemed to adjust their predictions depending on the task's difficulty. An effect of the covariate (judgment score for the first task) was also found, *F*(2,102) = 27.60, *MSe* = 5.59, $p < .001$, $\eta^2_p = .23$. No other main effect or interaction reached significance, all *Fs*<1.66, *ps*>.19.

< Figure 2 about here >

Calibration score. The results revealed a trend toward a main effect of age, *F*(2,102) = 2.84, *MSe* = 0.03, $p = .06$, $\eta^2_p = .06$, suggesting that older participants made more calibrated predictions than younger participants. We also found a main effect of task difficulty, *F*(1,102) = 30.61, *MSe* = 0.03, $p < .001$, $\eta^2_p = .24$, indicating that children overestimated their performance for the hard task (Mean = .09) and underestimated their performance for the easy task (Mean = -.06). Moreover, a feedback x task difficulty interaction was found, *F*(1,102) = 12.52, *MSe* = 0.03, $p < .001$, $\eta^2_p = .12$, (see Table 4). More specifically, children in the feedback condition underestimated their memory performance for the easy task (Mean = -.13) and overestimated their memory performance for the hard task (Mean = .14). Participants in the no feedback condition showed good calibration for both the easy (Mean = .01) and the hard (Mean = .04) task. In absolute terms, then, children in the feedback condition produced less calibrated responses than children in the no feedback condition, $p = .002$. Once again, an effect of the covariate (judgment score for the first task) was found, $F(2,102) = 12.11$, $MSe = 0.03$, $p < .001$, $\eta^2_p = .11$. No other main effect or interaction reached significance, all $Fs<1$, $ps>.89$.

< Table 4 about here >

Correlation analyses. We investigated the relationship between children's susceptibility to anchoring and feedback effects. Only participants in the high or low anchor conditions who also received feedback were included in the following correlation analyses (n = 36). A positive correlation was highlighted between children's predisposition to adjust their prospective judgments on the basis of an anchor and their inclination to rely on feedback to make their memory predictions, $r = .41$, $p = .01$ (see Figure 3).

Lastly, we examined the relationship between children's susceptibility to anchoring and feedback effects and their chronological age. Neither of the two correlations was significant, suggesting that children's sensitivity to external cues did not vary with age.

< Figure 3 about here >

Discussion

Experiment 3 demonstrated that children in the feedback condition adjusted their prediction in the direction of the feedback provided, regardless of task difficulty. Specifically, participants generated a higher prospective judgment for the hard task (after receiving positive feedback for the easy task) than for the easy task (after receiving negative feedback for the hard task). On the other hand, children in the no feedback condition seemed to take into account the strength of the association between the items that composed each word pair in the memory task when making their decisions. These children produced lower judgments for the hard task than for the easy task. These findings are not consistent with the hypothesis that providing feedback helps children to realize that they have not estimated their memory performance accurately, which in turn is supposed to prompt them to invest more efforts in monitoring the quality of their memories. Within this framework, children in the feedback condition would have been expected to adjust their memory predictions as a function of task difficulty. Indeed, many studies have demonstrated that children are able to detect differences in the strength of the association between the two items in a pair, and then use this information to guide their metacognitive judgments (e.g., Koriat et al., 2009a); this is confirmed by the pattern of results obtained in the no feedback condition. Children in that condition were shown to take the difficulty of the task into account when making their memory predictions.

On the other hand, our results seem to be in line with the anchoring hypothesis. The fact that participants in the feedback condition adjusted their predictions in the direction of the feedback without considering the task's difficulty is coherent with the hypothesis that the external information provided by the feedback was used as an anchor for judgment. Moreover, these findings are strengthened by the positive correlation found between the two scores computed to assess participants' susceptibility to the anchoring and feedback effects. The latter result indicates that children who are more sensitive to the anchoring effect are also more sensitive to the feedback effect, suggesting that, in this experiment, both effects might be sustained by the same mechanisms and, thus, that feedback influences children's judgments through automatic rather than effortful metacognitive processes.

General Discussion

This study was designed to achieve three main goals: (a) determine whether metacognitive feedback allows young children to better predict their memory performance; (b) examine whether children rely on the Anchoring-and-Adjustment heuristic to guide their memory decisions; and (c) explore whether the feedback effect results from an anchoring effect. Three experiments were carried out for this purpose. Their results are discussed in relationship to each of our hypotheses in the following sections.

Anchoring Effect

The involvement of automatic inference rules in children's decision processes has recently been demonstrated for various kinds of metacognitive heuristics (Geurten, Lloyd & Willems, 2016; Geurten, Meulemans, & Willems, 2015; Geurten, Willems, Germain, & Meulemans, 2015; Geurten, Willems, & Meulemans, 2015; Koriat et al., 2009a, 2009b). However, while some of these studies showed that heuristic-based inferences can already influence children's decisions by the age of 4 (e.g., Geurten, Willems, & Meulemans, 2015), others did not detect any reliance on metacognitive heuristics before the age of 9 (e.g., Koriat et al., 2009a, 2009b). Moreover, despite its acknowledged influence on various numerical judgments, the Anchoring-and-Adjustment heuristic had never been studied in children under the age of 9 (Smith, 1999). Experiments 2 and 3 consistently indicated that all children – whatever their age – adjusted their memory predictions lower or higher as a function of the external anchor, demonstrating the involvement of the Anchoring-and-Adjustment heuristic in their decision-making. From a theoretical point of view, this finding is coherent with other studies that have recently established that the ability to implement automatic inference rules to regulate decision-making processes develops very early on, providing interesting information on the types of mechanisms that underlie young children's decision-making.

Feedback Effect

Three interesting findings about the effect of feedback on global prospective judgments were observed. Experiment 1 suggested that feedback helps even very young learners to correct their metacognitive judgments when it is a direct indicator of future memory performance. However, when feedback is not a relevant predictor of performance on a subsequent test (Experiment 3), we found no difference in the calibration of prospective judgments between the two feedback conditions. More specifically, the results of the third experiment indicated that participants who received feedback predicted that they would remember more items for the hard task than for the easy task, while participants who did not receive feedback demonstrated the reverse pattern. This leads to the counterintuitive finding that *absolute* prediction accuracy in the no feedback condition was higher than in the feedback condition. In addition, a positive relationship was highlighted between children's susceptibility to the anchoring and feedback effects.

As a whole, these results are in line with the anchoring hypothesis (Scheck & Nelson, 2005) but not with the hypothesis that providing feedback draws children's attention to the fact that they have over- or underestimated their performance, which in turn should cause them to invest more efforts in evaluating the quality of their internal processes. Indeed, in Experiment 3, if learners truly allocated more resources to evaluating which items they knew and which items they did not, differences in the number of strongly associated pairs in the two tasks should have been detected (Koriat et al., 2009a). However, our result can easily be explained within the framework of the Anchoring-and-Adjustment heuristic. As a reminder, the anchoring effect is assumed to occur because people incompletely adjust their judgment from an anchor toward the value they would have given in the absence of such external information (Tversky & Kahneman, 1974). If information provided by metacognitive feedback serves as an anchor for the participant's memory predictions, it could explain why differences in task difficulty were not used to guide prospective judgment. When the easy task was administered first, children performed well, so they received positive feedback. Because they did not adjust their predictions away from this high value, they predicted that they would perform better on the hard task than the children who received no feedback did. Similarly, when the hard task was given first, children performed poorly, so they received negative feedback. Because they based their predictions for the easy task on this low anchor, they expected to remember fewer items than children in the no feedback condition. In other words, we postulate that the feedback provided for the first memory task had a lasting anchoring effect on children's judgments for the second memory task and that this anchoring effect was so strong that participants overlooked the task's difficulty when making subsequent judgments. The positive correlation between the two susceptibility scores further confirms this postulate since it indicates that children who were sensitive to the feedback effect were also sensitive to the anchoring effect. In other words, the participants who adjusted their predictions the most based on the anchor in the first task also adjusted their predictions the most based on the feedback during the second task. These results suggest that the same processes could be involved in anchoring-based and feedback-based judgments. Other studies examining the direct link between anchoring and feedback effect and the variables that sustain them should, of course, be carried out to corroborate our results. Moreover, several limitations should be noted for this research. First, multiple studies have shown that numerical understanding differs in preschool years and in elementary school years (e.g., Dowker, 2008). This difference could have influenced how numerical judgments were made in our three age groups. Another limitation is the small number of subjects per cell. Indeed, a larger sample size could have possibly led to different interpretations of our effects. Nonetheless, our findings already provide interesting information on how numerical feedback can affect children's metacognitive judgments.

Theoretical and Practical Perspectives

Theoretically, our findings are consistent with Koriat's (2007) model, which assumed that people do not systematically base their metacognitive judgments on effortful monitoring processes. Rather, they frequently rely on automatic inference rules, such as the Anchoring-and-Adjustment heuristic, to make quick decisions even though these heuristic-based decisions are prone to bias (Besken & Mulligan, 2013; Tversky & Kahneman, 1974). Interestingly, from a developmental point of view, our results showed that children in all age groups were inclined to use feedback as an anchor for their prospective judgments, suggesting that even older children with more cognitive resources still preferred to make their memory judgments on the basis of salient external information rather than to implement effortful monitoring processes. These results are strengthened by the finding that children's susceptibility to anchoring and feedback effects does not vary with age. Overall, these results suggest that children prefer to base their metacognitive decisions on automatic inference and that providing feedback does not seem to induce them to engage in more effortful decision-making processes.

Moreover, our results have practical implications. Feedback is usually considered to inform learners of what they know and what they do not know so that they will be able to efficiently regulate their future performance (for a review, see Kornell & Bjork, 2007). Based on this assumption, feedback is often given after learners have completed a task, particularly in educational settings. In our study, we found that even young children are able to use feedback to regulate their metacognitive decisions. However, receiving feedback does not always have a positive impact on children's judgments. Specifically, when we varied the level of difficulty between the task on which the feedback was given and the task on which the feedback had to be used, we found that children who received metacognitive feedback demonstrated poorer calibration than children who did not. Thus, it seems that, in some cases, feedback can impair the calibration of children's metacognitive judgment rather than improving it, suggesting that free rein should be given to young learners when they need to judge the quality of their memory. At the very least, the results of the present study indicate that feedback about the accuracy of metacognitive judgments should be used with caution. According to studies of feedback effects on cognitive performance, people do not react to all types of feedback in the same way (e.g., Allwood, Jonsson, & Granhag, 2005; Pashler, Cepeda, Wixted, & Rohrer, 2005). Considering the proven influence of metacognition on both cognitive and academic performance, future studies (e.g., using other types of metacognitive judgments and other types of feedback) should be conducted so we can better understand when and how children can take advantage of metacognitive feedback.

References

Allwood, C. M., Jonsson, A.-C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology, 19*, 331–344. doi:10.1002/acp.1071

Besken, M., & Mulligan, N. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory and Cognition, 41*, 897–903. doi:10.3758/s13421-013-0307-8

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 918–928. doi:10.1037/0278-7393.34.4.918

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Developmental Science, 11*, 650–654. doi:10.1111/j.1467-7687.2008.00713.x

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*, 597–607. doi:10.1037/0882-7974.13.4.597

Efklides, A., & Dina, F. (2004). Feedback from one's self and from the others: Their effect on affect. *Hellenic Journal of Psychology, 1*, 179–202.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin, 32*, 188–200. doi:10.1177/0146167205282152

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17*, 311–318. doi:10.1111/j.1467-9280.2006.01704.x

Everson, H., & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science, 26*, 65–79. doi:10.1023/A:1003040130125

Geurten, M., Lloyd, M.E., & Willems, S. (2016). Hearing “quack” and remembering a duck: Evidence for fluency attribution in young children. *Child Development*.

Geurten, M., Meulemans, T., & Willems, S. (2015). Memorability in context: A heuristic story. *Experimental Psychology, 62,* 306–319*.* doi:10.1027/1618-3169/a000300

Geurten, M., Willems, S., Germain, S., & Meulemans, T. (2015). Less is more: The availability heuristic in early childhood. *British Journal of Developmental Psychology, 33*, 405–410. doi:10.1111/bjdp.12114

Geurten, M., Willems, S., & Meulemans, T. (2015). Beyond the experience: Detection of metamemorial regularities. *Consciousness and Cognition, 33*, 16–23. doi:10.1016/j.concog.2014.11.009

Hacker, D., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning, 3*, 101–121. doi:10.1007/s11409-008-9021-5

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284. doi:10.1037/0033-2909.119.2.254

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., Ackerman, R., Lockl, K., & Schneider, W. (2009a). The easily learned, easily remembered heuristic in children. *Cognitive Development, 24*, 169–182. doi:10.1016/j.cogdev.2009.01.001

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009b). The memorizing effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology, 102*, 265–279. doi:10.1016/j.jecp.2008.10.005

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. doi:10.1006/jecp.2000.2612

Kornell, N., & Bjork, R. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin and Review, 14*, 219–224. doi:10.3758/BF03194055

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 609–622. doi:10.1037/0278-7393.32.3.609

Kornell, N., & Rhodes, M. G. (2013). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied, 19*, 1–13. doi:10.1037/a0032147

Kvavilashvili, L., & Ford, R. M. (2014). Metamemory prediction accuracy for simple prospective and retrospective memory tasks in 5-year-old children. *Journal of Experimental Child Psychology, 127*, 65–81. doi:10.1016/j.jecp.2014.01.014

Lipko, A. R., Dunlosky, J., Lipowski, S. L., & Merriman, W. E. (2011). Young children are not underconfident with practice: The benefit of ignoring a fallible memory heuristic. *Journal of Cognition and Development, 13*, 174–188. doi:10.1080/15248372.2011.577760

Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology, 103*, 152–166. doi:10.1016/j.jecp.2008.10.002

McElroy, T., & Dowd, K. (2007). Susceptibility to anchoring effects: How openness-to-experience influences responses to anchoring cues. *Judgment and Decision Making, 2*, 48–53.

Metcalfe, J., & Finn, B. (2011). People's hypercorrection of high-confidence errors: Did they know it all along? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 437–448. doi:10.1037/a0021962

Metcalfe, J., & Finn, B. (2012). Hypercorrection of high confidence errors in children. *Learning and Instruction, 22*, 253–261. doi:http://dx.doi.org/10.1016/j.learninstruc.2011.10.004

Miller, T., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning, 6*, 303–314. doi:10.1007/s11409-011-9083-7

Nelson, D., McEvoy, C., & Schreiber, T. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, and Computers, 36*, 402–407. doi:10.3758/BF03195588

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: The MIT Press.

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes, 39*, 84–97. doi:10.1016/0749-5978(87)90046-X

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3–8. doi:10.1037/0278-7393.31.1.3

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33–40. doi:10.1037/0022-0663.82.1.33

Radeau, M., Mousty, P., & Content, A. (1990). Brulex. Une base de données lexicales informatisée pour le français écrit et parlé. *L'année psychologique*, *90,* 551–566.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8*, 364–382. doi:10.1207/s15327957pspr0804_3

Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *British Journal of Educational Psychology, 79*, 749–767. doi:10.1348/978185409X429842

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*, 267–280. doi:10.1016/j.learninstruc.2010.07.004

Scheck, P., & Nelson, T. O. (2005). Lack of pervasiveness of the underconfidence-with-practice effect: Boundary conditions and an explanation via anchoring. *Journal of Experimental Psychology: General, 134*, 124–128. doi:10.1037/0096-3445.134.1.124

Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The adaptive nature of children's overestimation in a strategic memory task. *Cognitive Development, 22*, 197–212. doi:10.1016/j.cogdev.2006.10.001

Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on anchoring and adjustment: Do people adjust from provided anchors? *Journal of Personality and Social Psychology, 99*, 917–932. doi:10.1037/a0021540

Smith, H. D. (1999). Use of the anchoring and adjustment heuristic by children. *Current Psychology, 18*, 294–300. doi:10.1007/s12144-999-1004-4

Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 255–262. doi:10.1037/a0017892

Switzer, F. S., & Sniezek, J. A. (1991). Judgment processes in motivation: Anchoring and adjustment effects on judgment and behavior. *Organizational Behavior and Human Decision Processes, 49*, 208–229. doi:10.1016/0749-5978(91)90049-Y

Thiede, K. W. (1996). The relative importance of anticipated test format and anticipated test difficulty on performance. *The Quarterly Journal of Experimental Psychology Section A, 49*, 901–918. doi:10.1080/713755673

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131. doi:10.1126/science.185.4157.1124

Van Overschelde, J., & Nelson, T. (2006). Delayed judgments of learning cause both a decrease in absolute accuracy (calibration) and an increase in relative accuracy (resolution). *Memory and Cognition, 34*, 1527–1538. doi:10.3758/BF03195916

Wechsler, D. (2004). *Echelle d'intelligence de Wechsler pour la période préscolaire et primaire: WPPSI-III.* Paris: Les Editions du Centre de Psychologie Appliquée.

Wechsler, D. (2005). *Echelle d'intelligence de Wechsler pour enfants: WISC-IV.* Paris: Les Editions du Centre de Psychologie Appliquée.

Wright, W. F., & Anderson, U. (1989). Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Processes, 44*, 68–82. doi:10.1016/0749-5978(89)90035-6

Yussen, S. R., & Levy, V. M. (1975). Developmental changes in predicting one's own span of short-term memory. *Journal of Experimental Child Psychology, 19*, 502–508. doi:10.1016/0022-0965(75)90079-X

Appendix

Lists of Stimuli Used in Experiment 3

Easy list	**Hard list**
Collier – Bague (Necklace – Ring)	Peigne – Cheveux (Comb – Hair)
Hiver – Eté (Winter – Summer)	Jour – Nuit (Day – Night)
Ecole – Maîtresse (School – Teacher)	Pomme – Banane (Apple – Banana)
Fenêtre – Porte (Window – Door)	Table – Chaise (Table – Chair)
Fermier – Tracteur (Farmer – Tractor)	Sel – Sucre (Salt – Sugar)
Silence – Bruit (Silence – Noise)	Chambre – Forêt (Room – Forest)
Tigre – Lion (Tiger – Lion)	Assiette – Crayon (Plate – Pencil)
Boule – Glace (Scoop – Ice cream)	Cochon – Plage (Pig – Beach)
Roi – Reine (King – Queen)	Ballon – Rêve (Balloon – Dream)
Noir – Blanc (Black – White)	Muguet – Papier (Lily of the valley – Paper)
Lune – Jus (Moon – Juice)	Guitare – Veste (Guitar – Jacket)
Fantôme – Pelle (Ghost – Spade)	Nuage – Clef (Cloud – Key)
Feuille – Marmite (Leaf – Pot)	Ciseaux – Lampe (Scissors – Lamp)
Tambour – Caillou (Drum – Pebble)	Eléphant – Gomme (Elephant – Eraser)
Haricot – Poupée (Bean – Doll)	Tête – Livre (Head – Book)

Table 1

Mean Number of Items Recalled, Number of Items Predicted, and Prediction Calibration for the Two Memory Tasks, Each Experimental Condition (Feedback/No Feedback), and Each Age Group (4-, 6-, or 8-year-olds) in Experiment 1

	Feedback (n = 24)				No Feedback (n = 24)			
	4 years	6 years	8 years	Total	4 years	6 years	8 years	Total
Recall (1)	1.88 (1.36)	3.62 (1.77)	4.00 (1.41)	3.17 (1.74)	2.12 (0.99)	3.75 (1.03)	4.00 (1.31)	3.29 (1.37)
Recall (2)	1.50 (1.20)	3.25 (1.58)	3.75 (1.16)	2.83 (1.61)	1.37 (1.30)	2.13 (1.73)	3.50 (1.51)	2.33 (1.71)
Prediction (1)	10.12 (4.82)	6.87 (3.36)	6.75 (2.25)	7.92 (3.82)	7.75 (5.26)	7.75 (2.49)	8.13 (2.47)	7.88 (3.49)
Prediction (2)	6.87 (4.82)	4.87 (2.23)	5.87 (1.25)	5.88 (3.13)	7.5 (4.50)	8.25 (2.37)	8.12 (1.81)	7.96 (3.00)
Calibration (1)	0.62 (0.35)	0.25 (0.27)	0.18 (0.13)	0.35 (0.32)	0.38 (0.37)	0.27 (0.15)	0.27 (0.19)	0.31 (0.25)
Calibration (2)	0.36 (0.38)	0.12 (0.17)	0.14 (0.14)	0.21 (0.27)	0.41 (0.31)	0.41 (0.22)	0.31 (0.16)	0.38 (0.23)

Table 2

Mean Number of Items Recalled, Number of Items Predicted, and Prediction Calibration for Each Experimental Condition (Low, High, or No Anchor) and Each Age Group (4-, 6-, or 8-year-olds) in Experiment 2

	Low (n = 15)				Ø (n = 15)				High (n = 15)			
	4 years	6 years	8 years	Total	4 years	6 years	8 years	Total	4 years	6 years	8 years	Total
Recall	2.00 (1.87)	3.60 (1.14)	5.40 (1.82)	3.67 (2.09)	3.60 (3.29)	4.60 (1.14)	5.80 (1.48)	4.67 (2.43)	2.60 (1.14)	6.20 (2.68)	5.20 (1.10)	4.67 (2.29)
Prediction	5.60 (2.97)	5.40 (2.79)	4.80 (1.92)	5.27 (2.43)	7.80 (4.09)	6.00 (2.34)	8.60 (2.30)	7.47 (3.02)	9.80 (3.11)	8.80 (1.92)	8.20 (2.05)	8.93 (2.34)
Calibration	0.24 (0.14)	0.12 (0.24)	0.04 (0.20)	0.11 (0.22)	0.28 (0.32)	0.09 (0.20)	0.18 (0.24)	0.19 (0.25)	0.48 (0.18)	-0.17 (0.15)	0.20 (0.13)	0.28 (0.20)

Table 3

Mean Number of Items Recalled, Number of Items Predicted, and Prediction Calibration for the First Memory Task as a Function of Experimental Condition (Low, High, or No Anchor) and Age Group (4-, 6-, or 8-year-olds) in Experiment 3

	Low (n = 36)				Ø (n=36)				High (n = 36)			
	4 years	6 years	8 years	Total	4 years	6 years	8 years	Total	4 years	6 years	8 years	Total
Recall (1)	4.58 (2.11)	5.58 (1.62)	7.50 (1.98)	5.89 (2.23)	3.58 (1.78)	6.58 (1.68)	7.17 (2.37)	5.78 (2.49)	4.17 (2.21)	6.67 (2.39)	6.67 (3.03)	5.83 (2.76)
Prediction (1)	3.17 (3.24)	3.58 (1.08)	4.17 (1.27)	3.64 (2.09)	3.41 (1.51)	5.08 (2.47)	6.67 (2.71)	5.05 (2.60)	8.25 (2.93)	7.42 (2.97)	7.83 (2.44)	7.83 (2.73)
Calibration (1)	–0.11 (0.27)	–0.15 (0.12)	–0.26 (0.20)	–0.17 (0.21)	–0.01 (0.16)	–0.12 (0.26)	–0.04 (0.27)	–0.06 (0.23)	0.31 (0.27)	0.06 (0.19)	0.09 (0.26)	0.15 (0.26)

Table 4

Mean Number of Items Recalled, Number of Items Predicted, and Prediction Calibration for the Two Memory Tasks, Each Experimental Condition (Feedback/No Feedback) and Each Age Group (4-, 6-, or 8-year-olds) as a Function of Task Difficulty (Easy/Hard) in Experiment 3

		Recall (1)	Recall (2)	Prediction (1)	Prediction (2)	Calibration (1)	Calibration (2)
		Easy	Hard	Easy	Hard	Easy	Hard
Feedback	4 years	6.11 (0.60)	3.22 (0.83)	5.22 (3.42)	4.56 (2.13)	–0.07 (0.26)	0.10 (0.14)
	6 years	8.00 (1.58)	5.11 (2.71)	5.44 (2.40)	7.44 (2.60)	–0.20 (0.18)	0.18 (0.19)
	8 years	8.56 (1.88)	6.33 (1.94)	5.67 (2.35)	8.22 (3.31)	–0.22 (0.19)	0.15 (0.17)
	Total	7.56 (1.76)	4.89 (2.31)	5.44 (2.67)	6.74 (3.07)	-0.16 (0.22)	0.14 (0.17)
No Feedback	4 years	4.67 (1.00)	2.89 (1.90)	4.00 (2.60)	3.22 (1.09)	–0.05 (0.23)	0.03 (0.18)
	6 years	7.11 (0.78)	4.22 (1.64)	6.78 (3.15)	5.56 (2.51)	–0.03 (0.26)	0.10 (0.25)
	8 years	9.11 (1.54)	5.78 (1.20)	7.11 (2.80)	5.78 (1.72)	–0.15 (0.28)	0.00 (0.13)
	Total	6.96 (2.16)	4.30 (1.96)	5.96 (3.09)	4.85 (2.14)	-0.07 (0.26)	0.04 (0.19)
		Recall (1)	Recall (2)	Prediction (1)	Prediction (2)	Calibration (1)	Calibration (2)
		Hard	Easy	Hard	Easy	Hard	Easy
Feedback	4 years	3.56 (1.59)	5.67 (1.32)	5.33 (4.30)	3.22 (1.39)	0.14 (0.33)	–0.19 (0.12)
	6 years	5.11 (1.62)	6.22 (1.48)	4.33 (1.87)	5.44 (2.46)	–0.06 (0.15)	–0.06 (0.18)
	8 years	5.33 (2.18)	7.67 (2.45)	5.33 (2.18)	5.78 (1.30)	0.00 (0.25)	–0.15 (0.13)
	Total	4.67 (1.92)	6.52 (1.95)	5.00 (2.91)	4.81 (2.08)	0.02 (0.26)	-0.13 (0.15)
No Feedback	4 years	2.11 (2.08)	5.22 (0.97)	5.21 (3.96)	5.78 (2.95)	0.24 (0.27)	0.04 (0.22)
	6 years	4.89 (1.69)	6.56 (1.33)	4.89 (3.22)	6.78 (3.11)	0.00 (0.21)	0.02 (0.27)
	8 years	5.44 (1.42)	8.00 (2.65)	6.78 (3.23)	7.56 (3.61)	0.10 (0.30)	–0.03 (0.17)
	Total	4.15 (2.25)	6.59 (2.08)	5.63 (3.45)	6.70 (3.20)	0.11 (0.27)	0.01 (0.22)

Figure Captions

Figure 1. Details of procedures for each of the three experiments.

Figure 2. Number of items predicted for the second memory task as a function of task difficulty (easy vs. hard) and experimental condition (feedback vs. no feedback) in Experiment 3. Error bars show the standard deviation of the mean.

Figure 3. Scatterplot for children's susceptibility to the anchoring effect as a function of their level of susceptibility to the feedback effect in Experiment 3. The Pearson's correlation is still significant when the three outliers are removed ($r = .34$ instead of .41).