



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Here is a hint! How children integrate reliable recommendations in their memory decisions

Diana Selmeczy*, Simona Ghetti

Department of Psychology, University of California, Davis, Davis, CA 95616, USA
Center for Mind and Brain, University of California, Davis, Davis, CA 95616, USA



ARTICLE INFO

Article history:

Received 1 March 2018

Revised 1 August 2018

Keywords:

Metamemory
Recognition memory
Recognition bias
Memory cues
Signal detection theory
Development
Information-seeking

ABSTRACT

Children's own memory is not the only reliable source of information about past events. Others may possess relevant knowledge, and children must learn to appropriately consider it in combination with their own memories. In the current study, we investigated 5-, 7-, and 9-year-olds' ($N = 72$) ability to incorporate probabilistically reliable (70% accurate) hints into their memory decisions. Results revealed that children across ages were appropriately sensitive to these cues without following them blindly and indiscriminately. Furthermore, individual differences in metamemory monitoring predicted overall accuracy improvements after receiving cues in 9-year-olds but not in 5- and 7-year-olds, revealing a developmental role of metamemory for discerning when cues are most informative or needed. Although 5-year-olds increased overall confidence in their memory after receiving invalid cues, they still preserved the capacity to monitor their memory in the face of inaccurate information. Overall, children were sensitive to reliable recommendations, but developing metacognitive mechanisms predicted judicious benefits from cues.

© 2018 Elsevier Inc. All rights reserved.

Introduction

Imagine Clark, a 9-year-old boy, who is picking out a book at a bookstore. He thinks he has not read the book before, but his mother suggests that, in fact, he has. Clark eventually realizes that he has read

* Corresponding author at: Department of Psychology, University of California, Davis, Davis, CA 95616, USA.

E-mail address: dselmeczy@ucdavis.edu (D. Selmeczy).

the book and decides to pick a different one instead. In this scenario, Clark made a decision on considering two sources of information, his own memory and his mother's suggestion, both of which may be generally reliable but not perfectly accurate. Children often make memory decisions based on multiple relatively reliable sources of information, yet little is known about how these sources are factored into the decision process.

In the current study, we investigated whether and how children incorporate probabilistically reliable cues into their memory decisions. For example, if Clark thinks that his mother always remembers his past reading choices, it would be reasonable for him to follow her suggestion. However, Clark might recognize that even his mother makes mistakes sometimes such as when she mistakes events in his life with those of his siblings. Therefore, he should evaluate the quality of his own memory to determine when he should rely on others for his choices. We hypothesize that metamemory monitoring, or the ability to accurately self-reflect on one's own memory accuracy (Nelson, 1990), plays a critical role in this process. In other words, when Clark feels unsure about the accuracy of his memory, he should rely less on his own memory and more on the recommendation; in contrast, when he is sure he remembers accurately, he should instead rely more on his own memory and less on the recommendation.

Research with adult populations has shown that individuals with better metamemory monitoring benefit most from reliable external information (Konkel, Selmeczy, & Dobbins, 2015; Selmeczy & Dobbins, 2013). In children, previous literature has established that metamemory monitoring skills are evident as young as 5 years such that confidence is higher following correct responses as opposed to incorrect responses (Hembacher & Ghetti, 2014; Roebers, Gelhaar, & Schneider, 2004). These abilities improve throughout middle childhood and adolescence (Fandakova et al., 2017) and increasingly support decision making (Destan, Hembacher, Ghetti, & Roebers, 2014; Hembacher & Ghetti, 2013; Koriat & Ackerman, 2010). We predicted that such age-related improvements in metamemory also underlie children's ability to incorporate reliable external information in their memory decisions.

Previous work has investigated children's use of information provided by others during learning (Mills, 2013). Children as young as preschool age differentiate between reliable and unreliable informants and are more likely to learn new word labels for new objects from a reliable informant (Koenig & Harris, 2005; Pasquini, Corriveau, Koenig, & Harris, 2007). However, in these previous studies, metamemory monitoring was not as critical for children's decision to follow the informant's suggestion because children lacked previous knowledge about the target content. Thus, the decision of following the informant did not require a comparison between the quality of their own knowledge and that of the informant.

Eyewitness memory research provides additional insight on how children's reports may be influenced by external information (Bruck & Ceci, 1999; Ceci & Bruck, 1993). Children's memory is typically assessed after they received misleading or neutral information about an experienced event. Results show that misleading questioning reduces accuracy relative to neutral or unbiased questioning and that this difference is larger in younger children compared with older children and adults (Gordon, Baker-Ward, & Ornstein, 2001; Paz-Alonso & Goodman, 2016; Roebers & Schneider, 2000). Furthermore, metamemory monitoring is also shown to be altered following misleading questioning in children but not in adults (Roebers & Howie, 2003; Roebers, 2002). Although some eyewitness memory studies have included leading questions that suggested the correct answer, the overall accuracy of all suggestions was 50% at best (Schwarz & Roebers, 2006). Thus, the suggestions were not reliable overall, and the effects of these questions on memory performance were rarely measured separately for accurate and neutral suggestions (Roebers & Schneider, 2005; Roebers, 2002). These experimental choices are reasonable given the focus of these previous studies on the impact of misinformation on children's accuracy. However, these experimental designs do not allow for drawing proper inferences on how children adjust their decision processes in response to reliable information. The paradigm used in the current research was designed, instead, to assess the effects of reliable hints on changes in decision processes and potential accuracy gains.

The current study

The goal of the current study was to examine how children incorporate probabilistically reliable information into their memory decisions and the metamemory mechanisms that support

developmental improvements in this ability. We provided participants with reliable cues (i.e., valid for 70% of trials and invalid for 30%) that indicated whether an upcoming recognition probe would be “likely old” or “likely new” (O'Connor, Han, & Dobbins, 2010; Selmeczy & Dobbins, 2013) and compared children's performance on these trials with that on uncued trials when no cues were provided. Children were informed that the cues were generally reliable, but not always accurate, prior to starting their memory test. Thus, children needed to understand the meaning of probabilistic information, and children as young as 4 or 5 years have been shown to understand likelihoods using verbal probability labels such as “definitely,” “might,” and “probably” (Lagattuta & Sayfan, 2011). In the current task, we explicitly indicated that cues would give the right answer most of the time and verified that children understood these instructions. This is in contrast to work on probability learning in which children come to learn unknown and varying probabilities (Schlottmann & Wilkening, 2011). Instead, the provision of individual cues on each trial allowed us to assess how children weighted this information known to be generally reliable, but not perfectly accurate, against their own memory.

We focused on 5, 7, and 9 years of age because memory and metamemory performance improve throughout middle childhood (Ghetti, Mirandola, Angelini, Cornoldi, & Ciaramelli, 2011; Hembacher & Ghetti, 2013). In addition, eyewitness memory literature suggests developmental decreases in conformity around this age range (Schwarz & Roebbers, 2006). We chose 5-year-olds as our youngest age group because they show evidence of metamemory monitoring, as indicated by higher confidence for accurate responses compared with inaccurate responses (Hembacher & Ghetti, 2014), and are able to complete recognition tests with a large number of trials (Ghetti, Qin, & Goodman, 2002).

We aimed at addressing several questions. Our first question was to determine whether children's decision processes are sensitive to probabilistically reliable cues. We did so by adopting a *signal detection* approach (Macmillan & Creelman, 2005); we measured how the response criterion (i.e., bias toward responding “old”) changed as a function of cue type (likely old vs. likely new). In signal detection theory, the response criterion represents the tendency (or lack thereof) to endorse the presence of a target (e.g., a percept, a memory); this tendency may be affected by a variety of factors (Macmillan & Creelman, 2005). For example, when the ratio of old versus new items is varied in a recognition memory test, individuals are more likely to endorse old items (i.e., liberal criterion) if they expect them to be frequent and are less likely to endorse old items (i.e., conservative criterion) if they expect them to be infrequent (Kantner & Lindsay, 2010; Rhodes & Jacoby, 2007). Using a similar cueing design as that of the current study, adults are shown to adopt a liberal criterion following a likely old cue and to adopt a conservative criterion following a likely new cue (Selmeczy & Dobbins, 2013). If children are able to process reliable cues, they should alter their decision style accordingly. These shifts are the prerequisite to optimize performance, but as discussed below, they are not hypothesized to be sufficient.

We hypothesized that even young children would demonstrate the ability to shift their decision biases but that younger children, as compared with older children, may over-rely on cues, consistent with eyewitness memory studies indicating stronger influence of suggestive information at younger ages (Roebbers, Schwarz, & Neumann, 2005; Schwarz & Roebbers, 2006). This would be demonstrated by a greater change in the response criterion between likely old and likely new cues in younger children relative to older children. However, unlike in eyewitness memory research, in the current study children were informed that the cue may occasionally be incorrect, and this knowledge may protect them from over-relying on cues.

Our second question was whether and how children judiciously used cues to improve overall accuracy. Although shifts in response bias are a prerequisite, they are not sufficient to achieve overall gains in performance accuracy (Selmeczy & Dobbins, 2013). For example, if Clark, the boy in our earlier example, always followed his mother's recommendations, he would demonstrate high sensitivity to cues, but he would be at risk of overall declines if his mother makes many mistakes. Ideally, then, Clark would follow his mother's recommendations only selectively and when he needs them the most, that is, when he thinks his own memory is inaccurate.

We propose that children's ability to introspect on the accuracy of their own memory, or metamemory monitoring, is critical to appropriately weight internal and external sources of information. In other words, optimal performance is achieved if individuals appropriately rely on external cues (which are often, but not always, accurate) when they feel uncertain about their own memory

and if they ignore external cues when they feel certain about their own memory. By appropriately weighting cues against one's own memory, individuals can effectively boost performance because they follow the reliable cues when they otherwise would have been guessing or had low performance. Evidence of this process is found in adults in the association between individual differences in metamemory monitoring (when no cues are available) and gains in accuracy (when cues are provided) (Konkel et al., 2015; Selmeczy & Dobbins, 2013). We examined the extent to which a similar relation is evident in children.

Metamemory monitoring, indicated by greater subjective confidence for accurate responses compared with inaccurate responses, has been linked to decision making in both adults (Hembacher & Ghetti, 2017; Koriat & Goldsmith, 1996) and children (Hembacher & Ghetti, 2013, 2014). Furthermore, metamemory monitoring develops throughout early and middle childhood (Destan et al., 2014; Fandakova et al., 2017; Ghetti, Hembacher, & Coughlin, 2013), with older children making increasingly finer distinctions in their memory states (Ghetti et al., 2011). In addition, increasing research demonstrates that younger children often fail to use metamemory monitoring to guide their decision making despite demonstrating metacognitive monitoring (Destan et al., 2014; Hembacher & Ghetti, 2013; Schneider & Lockl, 2008). This suggests that younger children might not be able to translate their metamemory monitoring judgments into appropriate decision-making strategies. Thus, we hypothesized that individual differences in metamemory monitoring would predict performance gains from cues to a greater extent in older children compared with younger children.

In addition to the main hypotheses, the current study also afforded us the opportunity to replicate previous findings. We aimed to confirm established results from previous literature (Hembacher & Ghetti, 2013, 2014), including reliable metamemory monitoring in the uncued condition across ages and age-related improvements in this ability. Finally, the current study offered an opportunity for an exploratory investigation of possible impairments in children's metamemory due to invalid cues based on evidence that misleading information impairs children's metamemory monitoring (Roebbers & Howie, 2003; Roebbers, 2002). Invalid cues may lead to inflated confidence for incorrect responses, and this may be particularly true earlier in development (Howie & Roebbers, 2007; Roebbers, 2002). However, previous research shows that even young children can follow instructions to ignore misleading information (Principe, Haines, Adkins, & Guiliano, 2010; Schaaf, Bederian Gardner, & Goodman, 2015), and they can correct their memory reports after being told that they had received misleading information (Ghetti & Castelli, 2006). Thus, it is possible that knowledge about cues' reliability may lessen the potentially negative impact of invalid cues on metamemory monitoring.

Method

Participants

A total of 72 children participated in this study: 24 5-year-olds ($M = 5.76$ years, $SD = 0.16$; 12 girls), 24 7-year-olds ($M = 7.39$ years, $SD = 0.29$; 12 girls), and 24 9-year-olds ($M = 9.44$ years, $SD = 0.32$; 12 girls). We conservatively assumed a medium effect size of $f = .19$ ($\eta_p^2 = .04$) based on previous studies demonstrating large age-related differences in metamemory monitoring and manipulations of this capacity (Koriat & Ackerman, 2010; Roebbers et al., 2004). In addition, adult literature shows a large effect size for cues (e.g., Selmeczy & Dobbins, 2013). A priori power analyses determined our sample size as appropriate to find a repeated-measures interaction between cue type (likely old or likely new) and age group (5-, 7-, or 9-year-olds), with 80% power for effect size of $f = .19$ using default values for the correlation among repeated measures ($r = .50$) and nonsphericity correction ($\epsilon = 1$) (Faul, Erdfelder, Lang, & Buchner, 2007). In addition, our sample size was powered at 80% to identify a medium to large main effect of age ($f = .32$, $\eta_p^2 = .09$) and a medium to small main effect of cue type ($f = .17$, $\eta_p^2 = .03$).

Among the 5-year-olds, 5 children contributed partial data, having completed at least half of the total trials; the overall findings were similar if analyses were conducted without these participants. Data from an additional 5 children (4 5-year-olds and 1 7-year-old) were collected and excluded due to near-chance performance during uncued recognition trials (percentage correct <55%). In total,

47 children were Caucasian/White, 4 were Asian, 1 was Native Hawaiian or Other Pacific Islander, 16 were mixed ethnicities, and 4 chose not to report their ethnicity. In addition, 12 participants reported being Hispanic or Latino. Families were recruited through flyers in the area of Davis, California, in the western United States, and most families were upper-middle socioeconomic status. All participants provided informed consent in accordance with the University of California, Davis institutional review board.

Materials

Stimuli

Stimuli included 192 black and white line drawings of animals and objects (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997). Word labels for the pictured drawings had an average age of acquisition of 4.5 years ($SD = 0.87$) (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Stimuli were divided into six sets of 32 items each. Two sets were used as study items and two sets as distracters, counter-balanced across participants. Because previous research has shown low false alarm rates for pictorial stimuli (Ghetti & Angelini, 2008), we used two measures to prevent ceiling performance; two stimulus sets were used as filler studied items (i.e., studied items that were not tested), and studied and novel items were selected to be semantically similar.

Overall, participants completed a total of 128 recognition test trials (64 old and 64 new), where 96 trials (48 old and 48 new) were preceded by a cue indicating whether the upcoming item would be “likely old” or “likely new.” Cues were approximately 70% reliable (i.e., correct 68 of 96 times). This probability was selected because it corresponds to a reliable level of cue accuracy, but the number of invalidly cued trials is sufficient to conduct meaningful analysis. This level of cue accuracy does not encourage a simple outsourcing strategy (i.e., follow the cue every time), allowing us to investigate how children weigh evidence and integrate cues. Finally, similar levels of reliability were used in previous work (Konkel et al., 2015). Specifically, the likely old cue was given to 34 old items (70.8% valid) and 14 new items (29.2% invalid). The likely new cue was similarly given to 34 new items (70.8% valid) and 14 old items (29.2% invalid). The remaining 32 trials were uncued recognition trials (16 old items and 16 new items) and were used to assess baseline performance. All trials were randomly intermixed for each participant. The task was presented using E-Prime software (Psychology Software Tools, Pittsburgh, PA, USA).

Vocabulary assessment

Standardized receptive vocabulary was assessed using the Peabody Picture Vocabulary Task (Dunn & Dunn, 2007). This assessment was included to verify that all our age groups were matched on a separate cognitive task and used pictorial materials (colored line drawings) that were similar to those used in the memory task.

Procedure

During study, participants were shown 128 items presented for 1250 ms each and made living/nonliving judgments. After a 10-min delay involving completion of age-appropriate connect-the-dots and maze activities, participants were trained on the recognition memory task. During training, participants completed two practice trials and were told to indicate “yes” for items previously presented (old responses) and “no” for items not previously presented (new responses). Participants then completed two additional practice trials and were asked to provide confidence ratings on a 3-point scale indicating whether they were “really sure,” “kind of sure,” or “not so sure” of their answer using similar instructions as previous research (Coughlin, Hembacher, Lyons, & Ghetti, 2015; Hembacher & Ghetti, 2014). Pictorial illustrations of a child making facial expressions with varying levels of confidence were displayed as anchors for children’s confidence responses. Participants were then introduced to the cues and were told that they would get some hints to help them during the memory task. Participants were given an audio cue using a human female voice and were shown a smiley face in the top right corner of the screen with either open or closed eyes. Participants were told, “When you see a face with open eyes, you will hear ‘It is likely that you have seen this one.’ This means that you

have likely seen that picture before. When you see a face with closed eyes, you will hear ‘It is likely that you have not seen this one.’ This means that you have likely not seen that picture before.” Thus, our likely old and likely new cues were described verbally and anchored pictorially with a smiley face with open and closed eyes, respectively. In addition, participants were explicitly informed that the cues were reliable with the following instructions: “The hints are there to help you, so they will be right most of the time. We still want you to pay attention and think for yourself, so every once in a while the hints will be wrong. But again, the hints are helpful, so your job is to try and use the hints to help you do better during the memory game.” Finally, participants were also told that there would be trials during which they would not receive any help, indicated by an empty circle and no audio, which we refer to as uncued trials (see Fig. 1). Immediately after these instructions, children indicated whether the hints were right “all of the time,” “most of the time,” or “only a little bit of the time” by pointing to a rectangle filled with varying levels of color on the screen. Participants were corrected and told that the cues were right most of the time if they did not provide the correct answer; they were never informed of the actual probability of the cues (i.e., 70%). Eight practice trials were completed using the cueing procedure (two uncued, two invalidly cued, and four validly cued trials) in which participants received feedback on their performance and explicit feedback on the cue’s validity. Each cue (likely old, likely new, or uncued empty circle) was always presented for 2500 ms prior to the appearance of the recognition probe and yes/no options. Recognition and confidence responses were self-paced.

After the practice phase, participants completed 64 test trials, followed by a short break. No feedback was provided about the accuracy of the individual cue or the individual memory response during testing. Participants were then again asked whether the cue was right all of the time, most of the time, or only a little bit of the time and were corrected if they did not provide the correct answer (i.e., most of the time); participants were not reminded of the reliability of the cues prior to this assessment. This was followed by the last 64 test trials. Finally, participants were reminded that the cue was right most

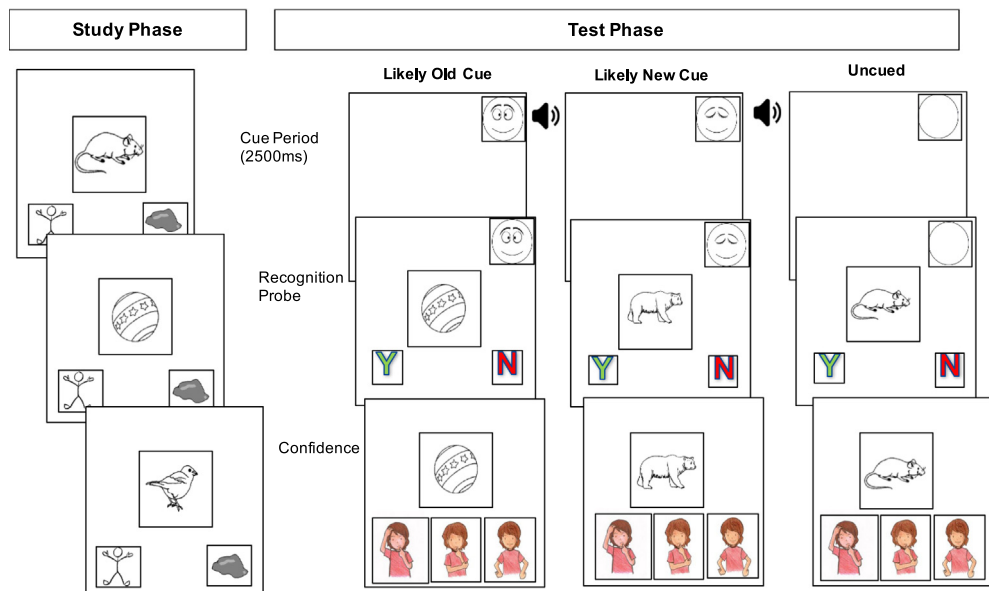


Fig. 1. Experimental design. During encoding, participants viewed pictures and made living/nonliving judgments anchored by an image of a person (living) and rock (nonliving). During test, participants were initially presented with a visual and auditory cue that indicated the likely status of the upcoming recognition probe, using a smiley face with open eyes (likely old) or closed eyes (likely new), or an uncued trial, using an empty circle. Cues were approximately 70% reliable. Participants then made a recognition judgment indicating whether they had seen the image before (Y, yes; N, no). Finally, participants indicated their confidence in their decision on a 3-point scale anchored by faces expressing different levels of confidence.

of the time and were asked to indicate how accurate they thought the cue was by pointing to buckets on the screen that were filled with color either 60%, 70%, 80%, or 90%. This question was asked to remind children that the cues were reliable (i.e., >50% correct) while providing us with the opportunity to assess graded differences in children's understanding that the cues were right most of the time. After finishing memory testing, participants completed the Peabody Picture Vocabulary Task (Dunn & Dunn, 2007).

To ensure that the effects of cues were not driven by differences in baseline recognition skill, we attempted to match uncued recognition performance across age groups using minor procedural adjustments that boosted performance for the 5-year-old group. The 5-year-olds completed the same number of trials as the older children; however, they completed two shorter study test cycles with no delay, and each item was presented for 2000 ms during encoding.

Results

Preliminary analyses

Receptive vocabulary

Preliminary analyses investigated performance on the Peabody Picture Vocabulary Task using standardized scores within each age group. Standardized scores did not significantly differ across age groups (5-year-olds: $M = 118.62$, $SD = 13.68$; 7-year-olds: $M = 119.58$, $SD = 12.46$; 9-year-olds: $M = 117.75$, $SD = 14.16$), $F(2, 69) = 0.11$, $p = .90$, $\eta_p^2 = .00$, suggesting typical developing receptive vocabulary across all age groups.

Accuracy for repeated test cycles




The 5-year-olds experienced two study test cycles instead of one longer study test cycle with all trials. Accuracy (d') during uncued trials did not significantly change between the first test cycle ($M = 1.49$, $SD = 0.75$) and second test cycle ($M = 1.66$, $SD = 0.85$) for 5-year-olds ($p = .57$), and therefore subsequent analysis includes data from both test cycles.

Cue reliability assessment

The proportions of children reporting each response option during cue reliability questioning are shown in Table 1. During the first assessment, immediately following the training instructions, children in each age group favored the "most of the time" response (71% of 5-year-olds, 71% of 7-year-olds, and 92% of 9-year-olds) and showed a distribution significantly different from chance (chi-square test, $ps < .001$). Furthermore, response and age group were not related (Fisher's exact test, $p > .27$). The second assessment of cue reliability occurred after the completion of the first half of the task, and children were not reminded about the reliability of the cues. During this second assessment, children favored the "most of the time" response (85% of 5-year-olds, 88% of 7-year-olds, and 83% of 9-year-olds), with distributions significantly different from chance in all age groups (chi-square test, $ps < .001$) and no relation between response and age group (Fisher's exact test, $p > .25$). The third assessment of cue reliability reminded children that the cues were reliable and asked them to indicate their assessment from 60%, 70%, 80%, and 90% response options. Children favored the 70% response option in each age group (47% of 5-year-olds, 54% of 7-year-olds, and 75% of 9-year-olds), with the distribution of counts being significantly different from chance in 9- and 7-year-olds (chi-square test, $ps < .006$) and marginally so in 5-year-olds (chi-square test, $p = .08$). Critically, the distribution of counts was not significantly different across age groups (Fisher's exact test, $p > .52$). Overall, these results suggest that all age groups appropriately and similarly understood that the cues provided the correct answer most of the time but not all of the time. Furthermore, despite never being given precise information about the actual accuracy level of the cues, the majority of children assessed the cues as being approximately 70% reliable, further demonstrating that they understood the probabilistic nature of the cues.

Table 1

Proportion of children selecting each response option for cue reliability ratings.

Age group	 All of the time	 Most of the time	 Little bit of the time
<i>First assessment during beginning of task</i>			
5-year-olds	.08	.71	.21
7-year-olds	.04	.71	.25
9-year-olds	.00	.92	.08
<i>Second assessment midway through task</i>			
5-year-olds	.10	.85	.05
7-year-olds	.00	.88	.12
9-year-olds	.00	.83	.17

Sensitivity to cues

To investigate age-related differences in sensitivity to cues, we first examined whether decision criteria appropriately shifted such that “old” response biases changed as a function of cue direction (see Fig. 2A). We predicted that all children would shift their decision criteria to cues but that younger children would over-rely on cues demonstrated by larger shifts. The criterion was calculated using signal detection measure C (Macmillan & Creelman, 2005), with zero representing unbiased responding, positive values indicating a conservative bias or tendency toward “new” responses, and negative values indicating a liberal bias or tendency toward “old” responses. We conducted a 3 (Age Group: 5-, 7-, or 9-year-olds) \times 3 (Cue Direction: likely old, likely new, or uncued) mixed analysis of variance (ANOVA) with cue direction varied within participants. We found a main effect of age, $F(2, 69) = 4.65$, $p = .01$, $\eta_p^2 = .12$, such that 9-year-olds exhibited less biased responding ($M = .14$, $SD = .34$) compared with 7-year-olds ($M = .44$, $SD = .31$, $p = .002$) but not 5-year-olds ($M = .28$, $SD = .37$, $p = .16$); the 5- and 7-year-olds did not differ from each other ($p = .13$). Across all age groups, we note that children's average criterion was significantly above zero ($p < .001$), indicating that children were generally conservative in their responding.

The main effect of cue type was also significant, $F(2, 138) = 17.04$, $p < .001$, $\eta_p^2 = .20$, such that relative to uncued trials ($M = .27$, $SD = .43$) participants favored “old” responses significantly more (i.e., lower C) under likely old cues ($M = .15$, $SD = .41$, $p = .005$) and significantly less (i.e., higher C) under likely new cues ($M = .43$, $SD = .44$, $p = .002$). In addition, both cued conditions affected responses similarly relative to uncued trials such that similar size shifts occurred toward likely new cues ($M = .16$, $SD = .41$) and likely old cues ($M = .12$, $SD = .36$, $p = .66$). The interaction with age was not significant ($p = .84$). These results suggest that children across age groups appropriately shifted their decision criteria in the direction of the cues and that, contrary to our hypothesis, this effect was similar across age groups. We note that children did not fully outsource to the cues because criterion values did not reach floor or ceiling (i.e., hit and false alarm rates were <1 and >0 , $ps < .001$) (Table 2). That is, if children fully outsourced to the cues, we would anticipate both hit and false alarm rates to be equal to 1 under a likely old cue (i.e., always respond “old”) and 0 under a likely new cue (i.e., always respond “new”).

A complementary way to assess sensitivity to cues is the examination of accuracy changes as a function of cue validity (i.e., whether or not the cue was accurate) regardless of direction. Thus, we entered discrimination accuracy (d') (Macmillan & Creelman, 2005) as the dependent measure in a 3 (Age Group: 5-, 7-, or 9-year-olds) \times 3 (Cue Validity: valid, invalid, or uncued) mixed ANOVA with cue validity varied within participants (see Fig. 2B). The main effect of age was marginally significant, $F(2, 69) = 2.82$, $p = .07$, $\eta_p^2 = .08$, with 9-year-olds ($M = 1.97$, $SD = 0.69$) exhibiting greater accuracy than 5-year-olds ($M = 1.58$, $SD = 0.42$, $p = .02$) but not 7-year-olds ($M = 1.73$, $SD = 0.59$, $p = .21$), who did not differ from each other ($p = .30$). The main effect of cue type was significant, $F(2, 138) = 19.11$, $p < .001$, $\eta_p^2 = .22$, such that relative to uncued trials ($M = 1.82$, $SD = 0.65$), accuracy significantly improved under

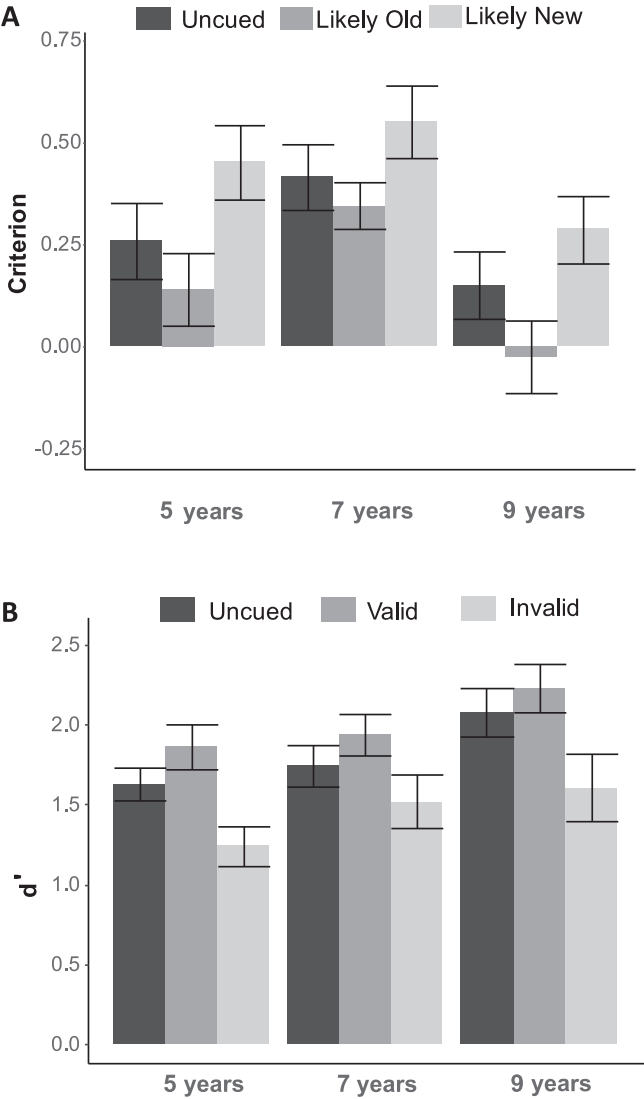


Fig. 2. (A) Criterion (*C*) as a function of cue direction. Smaller or negative values indicate favoring an “old” response, and larger or positive values indicate favoring a “new” response. (B) Accuracy (*d'*) as a function of cue validity. Error bars represent ±1 standard error.

valid cues ($M = 2.01$, $SD = 0.70$, $p = .02$) and declined under invalid cues ($M = 1.46$, $SD = 0.85$, $p < .001$). The interaction with age was not significant ($p = .79$). Thus, children were sensitive to cues such that performance increased and declined with cue validity, and this occurred to a similar extent across age groups.

Metamemory monitoring

Successful metamemory monitoring is demonstrated by higher confidence following correct responses as opposed to incorrect responses. We predicted age-related improvements in metamemory

Table 2
Mean hit and false alarm rates as a function of cues.

	Hit rate	False alarm rate
5-year-olds		
Uncued	.69 (.18)	.16 (.11)
Likely old	.72 (.16)	.19 (.15)
Likely new	.61 (.21)	.14 (.09)
7-year-olds		
Uncued	.66 (.17)	.12 (.11)
Likely old	.67 (.14)	.13 (.10)
Likely new	.61 (.21)	.09 (.07)
9-year-olds		
Uncued	.78 (.14)	.14 (.14)
Likely old	.80 (.12)	.22 (.20)
Likely new	.73 (.19)	.13 (.11)

Note. Standard deviations are in parentheses.

monitoring with age. In addition, we examined a secondary exploratory question and tested whether metamemory monitoring would decline under invalid cues. Confidence ratings were analyzed using a 3 (Age Group: 5-, 7-, or 9-year-olds) \times 2 (Response Accuracy: correct or incorrect) \times 3 (Cue Validity: uncued, valid, or invalid) mixed ANOVA with response accuracy and cue validity varied within participants (see Fig. 3).

Results revealed a main effect of age, $F(2, 68) = 3.54$, $p = .04$, $\eta_p^2 = .09$, such that 9-year-olds ($M = 1.29$, $SD = 0.24$) had lower overall confidence than 7-year-olds ($M = 1.50$, $SD = 0.34$, $p = .02$) and 5-year-olds ($M = 1.52$, $SD = 0.39$, $p = .02$). There was a main effect of response accuracy, $F(1, 68) = 83.36$, $p < .001$, $\eta_p^2 = .55$, such that confidence was higher following correct responses ($M = 1.59$,

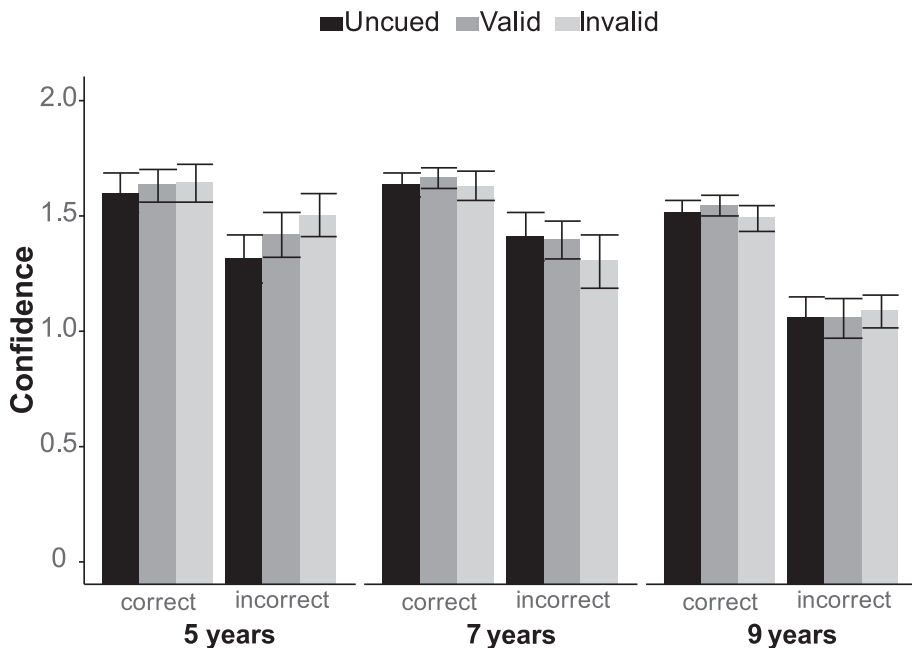


Fig. 3. Mean confidence for correct and incorrect responses as a function of cue validity. Confidence was coded as 2 ("really sure"), 1 ("kind of sure"), or 0 ("not so sure"). Error bars represent ± 1 standard error.

$SD = 0.30$) than following incorrect responses ($M = 1.28$, $SD = 0.43$, $p < .001$), and this was true in each age group separately ($ps < .001$), demonstrating that all age groups monitored their memory accuracy.

However, these effects were qualified by a significant Age Group \times Response Accuracy interaction, $F(2, 68) = 4.12$, $p = .02$, $\eta_p^2 = .11$, such that no age difference was observed among correct responses, $F(2, 68) = 1.27$, $p = .29$, $\eta_p^2 = .04$ (5-year-olds: $M = 1.62$, $SD = 0.38$; 7-year-olds: $M = 1.64$, $SD = 0.25$; 9-year-olds: $M = 1.51$, $SD = 0.23$), but 9-year-olds were significantly less confident in their incorrect responses ($M = 1.07$, $SD = 0.34$) than were 7-year-olds ($M = 1.37$, $SD = 0.45$, $p = .01$) and 5-year-olds ($M = 1.41$, $SD = 0.43$, $p = .004$), who did not differ from each other ($p = .74$). Thus, consistent with our predictions, typical developmental differences were observed in metamemory monitoring, and they were driven primarily by decreased incorrect confidence in older children.

In addition, there was significant Age Group \times Cue Validity interaction, $F(4, 136) = 2.60$, $p = .04$, $\eta_p^2 = .07$. In 5-year-olds, overall confidence was significantly higher under invalid cues ($M = 1.57$, $SD = 0.40$, $p = .01$) and numerically higher under valid cues ($M = 1.52$, $SD = 0.39$, $p = .11$) compared with uncued trials ($M = 1.45$, $SD = 0.40$); valid and invalid cueing conditions marginally differed from each other ($p = .10$). No such differences occurred in the other age groups ($ps > .24$). No remaining interactions or main effects were significant ($ps > .42$). To ensure that our findings were not driven by differences in memory performance, we included uncued memory accuracy (mean centered uncued d') as a covariate and confirmed Age Group \times Response Accuracy interaction, $F(2, 67) = 3.31$, $p = .04$, $\eta_p^2 = .09$, and Age Group \times Cue Validity interaction, $F(4, 134) = 2.38$, $p = .05$, $\eta_p^2 = .07$.

In contrast to previous research with misleading information (Roebers, 2002), the current exploratory analysis demonstrated that invalid cues did not decrease metamemory monitoring (Accuracy \times Cue Validity, $p = .76$); confidence remained higher for correct responses relative to incorrect responses similarly during all cueing conditions, even during invalid cues in all age groups ($ps < .05$). Instead, we found that younger children inappropriately increased overall confidence with cues, particularly when the cues were invalid. This finding was also confirmed when we tested 5-year-olds separately; we found a significant effect of cue validity, $F(2, 46) = 4.85$, $p = .01$, $\eta_p^2 = .17$, but no interaction between cue validity and response accuracy, $F(2, 46) = 1.12$, $p = .31$, $\eta_p^2 = .05$, suggesting that cues increased both correct and incorrect responses similarly. The lack of an interaction in 5-year-olds is somewhat surprising; if young children trusted cues generally more, their confidence should be expected to increase when the cue provides confirmatory evidence to children's intended responses (i.e., correct responses during valid cues and incorrect responses during invalid cues). For incorrect responses, confidence was indeed higher under invalid cues ($M = 1.50$, $SD = 0.47$, $p = .03$), but not under valid cues ($M = 1.42$, $SD = 0.47$, $p = .23$), relative to uncued trials ($M = 1.31$, $SD = 0.51$). For correct responses, confidence was similar during valid cues ($M = 1.63$, $SD = 0.35$, $p = .19$) and invalid cues ($M = 1.64$, $SD = 0.40$, $p = .23$) relative to uncued trials ($M = 1.60$, $SD = 0.42$). Thus, 5-year-olds seemingly gained additional confidence from invalid cues when they were the most uncertain (during incorrect responses) and the cues provided some confirmation for their incorrect selections.

Metamemory monitoring and cueing benefit

The group-level analysis revealed that all age groups were sensitive to cues, demonstrating that cues influenced children's decision processes. In the following analysis, we examine the factors that may contribute to judicious benefit from the cues. We predicted that accuracy benefits from the cues would be related to metamemory monitoring and that this relation would be stronger in older children relative to younger children. To investigate this question, we took an individual difference approach and examined whether those children with the highest metamemory monitoring ability were also the children who benefitted the most from cues.

For each participant, cueing benefit was defined as the difference in d' between cued and uncued conditions (i.e., cued d' minus uncued d'), where cued d' was calculated as overall accuracy across all cued trials regardless of validity. Metamemory monitoring was defined as mean confidence for correct trials minus mean confidence for incorrect trials in the uncued condition. Although metamemory monitoring was not significantly influenced by cues, we restricted our measure to the uncued condition to estimate a pure measure of metamemory monitoring ability not altered by cues and to remain consistent with work in adult and older adult populations (Konkel et al., 2015).

Table 3

Regression results predicting cueing benefit.

	<i>b</i>	SE <i>b</i>	<i>B</i>	<i>t</i> statistic	<i>p</i> value
(Intercept)	.56	.24	–	2.34	.02 [†]
Uncued <i>d'</i>	–.52	.10	–.56	–5.39	<.001 ^{***}
Metamemory monitoring	.98	.26	.72	3.83	<.001 ^{***}
Age: 7 years	.38	.20	–	1.94	.06
Age: 5 years	.22	.20	–	1.09	.28
Metamemory Monitoring × Age: 7-year-olds	–.85	.36	–.63	–2.39	.02 [†]
Metamemory Monitoring × Age: 5-year-olds	–.74	.34	–.54	–2.19	.03 [†]

Note. β represents regression weights for standardized continuous predictors.

[†] $p < .05$, ^{**} $p < .01$, ^{***} $p < .001$.

Simultaneous multiple regressions analysis was used to predict cueing benefit with the following predictors: metamemory monitoring, age (dummy coded relative to 9-year-olds), the interaction between age and metamemory monitoring, and uncued *d'*, which accounts for differences in uncued accuracy. This regression approach is very similar to the approaches previously used in adults (Konkel et al., 2015; Selmeczy & Dobbins, 2013). See Table 3 for the regression results.

We found that metamemory monitoring was a significant positive predictor of cueing benefit ($b = .98$, $p < .001$). However, this relation was fully qualified by an interaction between age and metamemory monitoring such that 5- and 7-year-olds showed a significantly weaker relation than did 9-year-olds ($b = -.74$, $p = .03$ and $b = -.85$, $p = .02$, respectively); the 5- and 7-year-olds did not differ from each other ($b = .12$, $p = .73$). To illustrate the interaction, we depict the relation between metamemory monitoring and improvements in *d'* separately in each age group (see Fig. 4). Results also revealed a significant effect of uncued *d'* ($b = -.52$, $p = .001$), demonstrating that those children with lower uncued performance experienced greater gains in performance. In summary, these results show that metamemory monitoring is an important predictor in cueing benefit; however, this relation does not emerge until later during middle childhood.

Finally, to reassure that metamemory monitoring was specifically related to judicious cueing benefit (i.e., improvements in *d'* from uncued to cued conditions) and not to sensitivity to cues (i.e., shifting in criterion between likely old and likely new cues), we conducted the same regression predicting shifts in criterion measured as likely new cued *C* minus likely old cued *C*. Whereas uncued *d'* was a marginally negative predictor ($\beta = -.16$, $p = .06$), suggesting that those with lower uncued *d'* experienced larger criterion shifts, no other effect was significant ($ps > .18$). These findings underscore that simply being sensitive to cues is not associated with metamemory monitoring; instead, metamemory monitoring is particularly important to achieve judicious performance benefits from cues.

Discussion

Children are routinely tasked with discerning when to trust others and when to trust themselves, a task that requires accurate metamemory monitoring. When we do not remember relevant information or we are uncertain about it, it is best to prioritize external information. In contrast, when we are confident in our memory, we should trust it if we have reason to believe that external information is not always accurate. In the current study, we asked how children engage in this assessment process.

We investigated whether children could adaptively incorporate probabilistically reliable cues into their memory decisions. Results revealed that children as young as 5 years could appropriately bias their responses based on cues without blindly relying on them. Critically, we predicted that developmental improvements in metamemory would be linked to children's ability to benefit from cues. Metamemory monitoring was observed consistently across all ages and improved with age. Importantly, metamemory monitoring predicted judicious cueing improvements in accuracy in 9-year-olds.

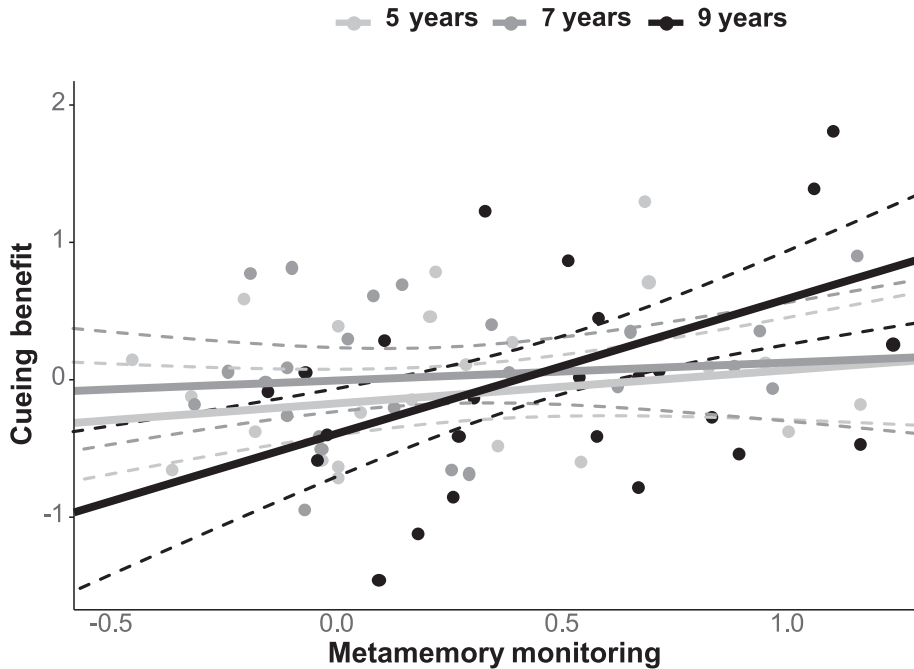


Fig. 4. Associations between cueing benefit and metamemory monitoring for each age group. Cueing benefit was calculated as cued d' minus uncued d' . Metamemory monitoring was calculated as average correct response confidence minus average incorrect response confidence during uncued trials. Solid lines represent the Metamemory Monitoring \times Age interaction, and dashed lines represent 95% confidence intervals estimated based on the entire regression model. The raw correlations between cueing benefit, with uncued d' partialled, and metamemory monitoring for each age group are provided for descriptive purposes (5-year-olds: $r = .27$, $p = .21$; 7-year-olds: $r = .16$, $p = .44$; 9-year-olds: $r = .57$, $p = .003$).

Children's sensitivity to cues and decision making

The ability to adaptively shift decision criteria has received considerable interest in the adult literature with researchers examining the limits of when criterion shifts occur and the underlying mechanisms that support this ability (Benjamin, Diaz, & Wee, 2009; Chua & Ahmed, 2016; Dobbins & Han, 2007; Han & Dobbins, 2009; Kantner & Lindsay, 2012). However, previous work has not examined criterion shifting in children (but see Berch & Evans, 1973) and the processes that may support its development. In the current study, we examined criterion shifts by providing participants with reliable probabilistic memory cues that suggested whether or not an item had been previously seen. We found that the ability to shift criteria is evident as early as 5 years of age, suggesting that even young children have the capacity to appropriately integrate external information into their memory decision making. This finding is consistent with the growing literature suggesting that shifting criteria may require less sophisticated cognitive control than previously thought, such that participants show shifts even when instructed to ignore cues (Selmecky & Dobbins, 2017) and older adults with known working memory deficits show similar size cueing shifts as young adults (Konkel et al., 2015). Although we did not detect developmental differences in this ability for children aged 5–9 years assuming a medium effect size, it is possible that a small effect size exists or developmental differences would emerge if more complex decision processes were necessary. For example, work in adults has shown that increasing the number of decision criteria that need to be considered throughout the task can decrease performance (Benjamin, Tullis, & Lee, 2013), and work in children has also suggested that older children are better at using multiple probabilistic cues during decision making (Betsch, Lang, Lehmann, & Axmann, 2014). In the current task, children were provided with a single reliable

source and did not need to discover the underlying probability of the cues. We showed that even 5-year-olds appropriately understood that probabilistically reliable cues would be correct most of the time but not all of the time, consistent with other work examining children's understanding of likelihoods (Lagattuta & Sayfan, 2011). Future work should assess whether developmental differences emerge when children must integrate recognition cues from multiple sources and/or differing reliabilities, which would require adopting and maintaining several different decision criteria and a more advanced understanding of probabilistic information.

We predicted that younger children would conform to the cues more than older children based on work in eyewitness memory paradigms showing that younger children are more likely to report misleading information than are older children (Roebbers & Schneider, 2005; Roebbers, 2002; Schwarz & Roebbers, 2006). Although we did not find developmental differences in children's use of reliable cues when we examined recognition responses, we did find developmental differences in confidence, consistent with younger children being at greater risk for conformity during incorrect responses. Several differences exist between the current paradigm and the paradigms used in eyewitness memory work. For example, in eyewitness memory work information is generally provided by confederates, and several social factors have been shown to influence the degree of conformity observed both in adults (Davis & Meade, 2013; Kieckhafer & Wright, 2015) and in children (Galindo & Harris, 2017; Schwarz & Roebbers, 2006). Based on our results, it is intriguing to consider that children may be better at incorporating generally reliable information than at resisting generally misleading information. However, the differences between our task and eyewitness work are too extensive to make direct comparisons. Future work should directly assess this hypothesis by manipulating the degree of cue reliability using a similar design as the current study and providing children with generally reliable versus generally unreliable or random recognition cues.

Metamemory monitoring and cue validity

Children were provided with generally reliable cues; by definition, some of them were invalid. Thus, the current study offered the opportunity to explore reported impairments in metamemory due to misleading information (Roebbers, 2002). We established that metamemory monitoring was present and robust across age groups and conditions even though 9-year-olds were better able to monitor their memories through decreased incorrect response confidence relative to 5- and 7-year-olds, consistent with previous research (Roebbers et al., 2004). Critically, metamemory monitoring has previously been found to be drastically reduced, if not eliminated, in children under misleading information (Roebbers, 2002), suggesting that children's confidence may be particularly sensitive to invalid cues. However, exploratory analysis demonstrated that there was no decline in metacognitive monitoring under invalid cueing. One potential reason for this finding is that participants knew that the cues were probabilistic and may occasionally be wrong, whereas in the eyewitness memory work children were not informed that misleading information may be provided (Roebbers & Howie, 2003; Roebbers, 2002). Therefore, knowledge regarding the reliability of external information may guard against the disruption of metamemory monitoring.

Although metamemory monitoring remained intact, our results showed that the overall confidence of 5-year-olds, but not of 9- and 7-year-olds, increased from cues, particularly under invalid cueing when confidence should actually decrease (Jaeger, Cox, & Dobbins, 2012). Because the cue was not on the screen during confidence judgments, it is unlikely that children reported higher confidence because they felt pressure to conform to a cue on the screen. Thus, in addition to developing metamemory monitoring in the absence of cues, young children also showed inflated confidence when they were provided with cues, suggesting that cues provided some confirmation for their selection. Specifically, the increase in confidence for incorrect responses under invalid cues is consistent with the idea that young children experienced the cues as a confirmation of their selected, but previously uncertain, wrong answers. These results highlight the importance of examining the development of metamemory under varying circumstances. Together, our results and those from previous research suggest that two critical dimensions play a role: the accuracy of the available information, given that children are shown to be sensitive to informants' quality (Pasquini et al., 2007; Schwarz & Roebbers, 2006), and the degree to which suggestive techniques are employed, given that

metacognitive monitoring is decreased with misleading questions (Roebers & Howie, 2003; Roebers, 2002) but not when children are informed that hints are not always accurate as in the current research. Future research manipulating these factors within one paradigm might be informative.

Finally, adults have been found to gain confidence from valid cues (Jaeger et al., 2012), suggesting that they gain confirmatory evidence from valid cues during correct responses (in addition to judiciously following the cues when they are uncertain (Selmezy & Dobbins, 2013)). We did not find any evidence of this pattern in older children. Future research should ascertain whether this is due to children's difficulty in considering converging evidence once children established that they have answers. Protracted development beyond 9 years of age is consistent with recent work that demonstrates some improvements in metacognitive abilities into early adolescence (Fandakova et al., 2017; Weil et al., 2013).

Metamemory monitoring and judicious incorporation of cues

Cue sensitivity is a prerequisite to judicious decision making but is not sufficient. Because the cues are occasionally invalid, children must evaluate their own memory signals and determine whether it is appropriate to rely on the cues because probabilistically reliable cues may help performance when children's memory is inaccurate but may potentially hurt performance when their memory is successfully retrieved. Children with better metamemory monitoring may be better at gauging when their memory is least likely to be accurate, thereby selectively following the cues when they need them the most. This prediction has been corroborated with younger adults (Selmezy & Dobbins, 2013) and older adults (Konkel et al., 2015). The current study joins this body of work in showing that individual differences in metamemory monitoring significantly predicted cueing benefit in 9-year-olds. In addition, it is important to note that no such correlation emerged with criterion shifting in children, suggesting that metamemory monitoring is not critical for sensitivity to external cues. Instead, metamemory monitoring emerged as a key predictor of gains in performance in older children. We note that our multiple regression analysis includes both uncued memory performance and metamemory monitoring as predictors, accounting for any age-related differences in these skills. Furthermore, to examine monitoring in the absence of cue influence, we estimated metamemory monitoring from confidence during uncued trials (i.e., 32 trials), which occurred less frequently than cued trials. This number of trials was selected in order to keep the length of the study appropriate for children while maximizing the number of invalid cues, which necessarily occurred less frequently than valid cues to achieve an overall reliability of 70%. In addition, our number of uncued trials is similar to or exceeds the number of unbiased trials used in eyewitness memory work (e.g., Roebers et al., 2005; Schwarz & Roebers, 2006). However, the possibility exists that our estimates of metamemory monitoring may be less reliable due to the relatively low number of uncued trials. We believe that this is unlikely because 7- and 9-year-olds' confidence was similar for uncued and valid trials even though valid trials occurred much more frequently (i.e., 68 trials). Furthermore, confidence during uncued trials yielded typical developmental improvements in metamemory (e.g., Destan et al., 2014), providing some reassurance about the reliability of our measure.

The finding that the relation between metamemory monitoring and cueing benefit was present only in older children is consistent with other literature suggesting that there is a developmental lag between the emergence of metamemory monitoring and appropriate use of this capacity in service of decision making such as selecting subsets of answers in anticipation of a reward (Destan et al., 2014; Hembacher & Ghetti, 2013). Critically, we add to this literature by demonstrating that effective metamemory monitoring also plays a role in how children integrate external information into memory decisions. Although additional research is necessary to explain the developmental lag between metacognitive monitoring and decision making, there is some evidence to suggest that these behaviors depend on partly different underlying processes. Metamemory monitoring may derive from partially implicit assessments of fluency (Koriat, 1997), and research shows that even young children respond to these factors in their metacognitive judgments (Geurten & Willems, 2016). In contrast, adaptive self-regulation of decisions may rely more heavily on executive control and strategy selection (Roebers, 2017), which develop extensively throughout middle childhood and adolescence (Fandakova et al., 2017, 2018). Thus, one possibility for the developmental lag between metamemory and decision

making may be that the underlying processes differ in their developmental trajectories. We recognize, however, that this developmental lag is not necessarily ubiquitous. Under certain circumstances, experience with decision making may lead to improvements in metacognitive monitoring (Koriat, Ackerman, Adiv, Lockl, & Schneider, 2014). Thus, future research should investigate whether the decision to actively seek hints, as opposed to responding to provided hints as in the current study, leads to a more careful evaluation of one's own memory, resulting in better metamemory monitoring estimates and positive relations between gains in accuracy and monitoring in young children.

Finally, related literature has examined how children learn to trust testimony from others when they learn something new (see Harris, Koenig, Corriveau, & Jaswal, 2018, for a review). A fruitful avenue of future research would be to assess the role of metacognition in children's ability to learn from testimony using designs in which children are required to evaluate their own performance while weighting information provided by one or multiple informants. We could then learn whether developmental differences in metacognitive monitoring play a role in children's ability not only to decide when to follow a recommendation, as documented here, but also to learn which informant is the most accurate or which informant should be trusted more.

Conclusions

Children's capacity to remember improves from infancy to adolescence. At the same time, the act of remembering is often shared with parents, siblings, and friends. Thus, children might need to entertain different versions of the same past more often than not and be tasked with evaluating the veracity of their own memories against information from these other sources. This ability to weigh evidence and regulate decision making is an important skill that children must learn as they grow to be independent thinkers and learners. The current study revealed that, 5- to 9-year-olds were similarly sensitive to external recommendations. However, developmental differences emerged with regard to the potential mechanisms that support the ability to judiciously benefit from cues such that metamemory monitoring predicted accuracy benefits from cues more strongly in 9-year-olds than in the younger age groups. Overall, our findings shed light on children's ability to regulate their memory decisions and the developing metacognitive mechanisms that support this process.

Acknowledgment

Support for this research was provided by the National Institute of Child Health and Human Development (F32HD084313).

References

- Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, 116, 84–115.
- Benjamin, A. S., Tullis, J. G., & Lee, J. H. (2013). Criterion noise in ratings-based recognition: Evidence from the effects of response scale length on recognition accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1601–1608.
- Berch, D. B., & Evans, R. C. (1973). Decision processes in children's recognition memory. *Journal of Experimental Child Psychology*, 16, 148–164.
- Betsch, T., Lang, A., Lehmann, A., & Axmann, J. M. (2014). Utilizing probabilities as decision weights in closed and open information boards: A comparison of children and adults. *Acta Psychologica*, 153, 74–86.
- Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of Psychology*, 50, 419–439.
- Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403–439.
- Chua, E. F., & Ahmed, R. (2016). Electrical stimulation of the dorsolateral prefrontal cortex improves memory monitoring. *Neuropsychologia*, 85, 74–79.
- Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science*, 18, 957–971.
- Cyrowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, 65, 171–237.
- Davis, S. D., & Meade, M. L. (2013). Both young and older adults discount suggestions from older adults on a social memory test. *Psychonomic Bulletin & Review*, 20, 760–765.
- Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5- to 7-year-old children. *Journal of Experimental Child Psychology*, 126, 213–228.

- Dobbins, I. G., & Han, S. (2007). What constitutes a model of item-based memory decisions? *Psychology of Learning and Motivation*, 48, 95–144.
- Dunn, D. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test manual* (4th ed.). Minneapolis, MN: NCS Pearson.
- Fandakova, Y., Bunge, S. A., Wendelken, C., Desautels, P., Hunter, L., Lee, J. K., & Ghetti, S. (2018). The importance of knowing when you don't remember: Neural signaling of retrieval failure predicts memory improvement over time. *Cerebral Cortex*, 28, 90–102.
- Fandakova, Y., Selmeczy, D., Leckey, S., Grimm, K. J., Wendelken, C., Bunge, S. A., & Ghetti, S. (2017). Changes in ventromedial prefrontal and insular cortex support the development of metamemory from childhood into adolescence. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 7582–7587.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Galindo, J. H., & Harris, P. L. (2017). Mother knows best? How children weigh their firsthand memories against their mothers' reports. *Cognitive Development*, 44, 69–84.
- Geurten, M., & Willems, S. (2016). Metacognition in early childhood: Fertile ground to understand memory development? *Child Development Perspectives*, 10, 263–268.
- Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development*, 79, 339–358.
- Ghetti, S., & Castelli, P. (2006). Developmental differences in false-event rejection: Effects of memorability-based warnings. *Memory*, 14, 762–776.
- Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. *Child Development Perspectives*, 7, 160–165.
- Ghetti, S., Mirandola, C., Angelini, L., Cornoldi, C., & Ciaramelli, E. (2011). Development of subjective recollection: Understanding of and introspection on memory states. *Child Development*, 82, 1954–1969.
- Ghetti, S., Qin, J., & Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective experience. *Developmental Psychology*, 38, 705–718.
- Gordon, B. N., Baker-Ward, L., & Ornstein, P. A. (2001). Children's testimony: A review of research on memory for past experiences. *Clinical Child and Family Psychology Review*, 4, 157–181.
- Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. *Psychonomic Bulletin & Review*, 16, 469–474.
- Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, 69, 251–273.
- Hembacher, E., & Ghetti, S. (2013). How to bet on a memory: Developmental linkages between subjective recollection and decision making. *Journal of Experimental Child Psychology*, 115, 436–452.
- Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological Science*, 25, 1768–1776.
- Hembacher, E., & Ghetti, S. (2017). Subjective experience guides betting decisions beyond accuracy: Evidence from a metamemory illusion. *Memory*, 25, 575–585.
- Howie, P., & Roebbers, C. M. (2007). Developmental progression in the confidence–accuracy relationship in event recall: Insights provided by a calibration perspective. *Applied Cognitive Psychology*, 21, 871–893.
- Jaeger, A., Cox, J., & Dobbins, I. G. (2012). Recognition confidence under violated and confirmed memory expectations. *Journal of Experimental Psychology: General*, 141, 282–301.
- Kantner, J., & Lindsay, D. S. (2010). Can corrective feedback improve recognition memory? *Memory & Cognition*, 38, 389–406.
- Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, 40, 1163–1177.
- Kieckhafer, J. M., & Wright, D. B. (2015). Likable co-witnesses increase eyewitness accuracy and decrease suggestibility. *Memory*, 23, 462–472.
- Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, 76, 1261–1277.
- Konkel, A., Selmeczy, D., & Dobbins, I. G. (2015). They can take a hint: Older adults effectively integrate memory cues during recognition. *Psychology and Aging*, 30, 781–794.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370.
- Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science*, 13, 441–453.
- Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General*, 143, 386–403.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490–517.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44, 978–990.
- Lagattuta, K. H., & Sayfan, L. (2011). Developmental changes in children's understanding of future likelihood and uncertainty. *Cognitive Development*, 26, 315–330.
- Macmillan, N., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology*, 49, 404–418.
- Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). San Diego: Academic Press.
- O'Connor, A. R., Han, S., & Dobbins, I. G. (2010). The inferior parietal lobule and recognition memory: Expectancy violation or successful retrieval? *Journal of Neuroscience*, 30, 2924–2934.

- Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, 43, 1216–1226.
- Paz-Alonso, P. M., & Goodman, G. S. (2016). Developmental differences across middle childhood in memory and suggestibility for negative and positive events. *Behavioral Sciences and the Law*, 34, 30–54.
- Principe, G. F., Haines, B., Adkins, A., & Guiliano, S. (2010). False rumors and true belief: Memory processes underlying children's errant reports of rumored events. *Journal of Experimental Child Psychology*, 107, 407–422.
- Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 305–320.
- Roebers, C. M. (2002). Confidence judgments in children's and adults' event recall and suggestibility. *Developmental Psychology*, 38, 1052–1067.
- Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, 45, 31–51.
- Roebers, C. M., Gelhaar, T., & Schneider, W. (2004). "It's magic!" The effects of presentation modality on children's event memory, suggestibility, and confidence judgments. *Journal of Experimental Child Psychology*, 87, 320–335.
- Roebers, C. M., Schwarz, S., & Neumann, R. (2005). Social influence and children's event recall and suggestibility. *European Journal of Developmental Psychology*, 2(1), 47–69.
- 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.
- Roebers, C. M., & Schneider, W. (2000). The impact of misleading questions on eyewitness memory in children and adults. *Applied Cognitive Psychology*, 14, 509–526.
- Roebers, C. M., & Schneider, W. (2005). The strategic regulation of children's memory performance and suggestibility. *Journal of Experimental Child Psychology*, 91, 24–44.
- Schaaf, J. M., Bederian Gardner, D., & Goodman, G. S. (2015). Gating out misinformation: Can young children follow instructions to ignore false information? *Behavioral Sciences and the Law*, 33, 390–406.
- Schlottmann, A., & Wilkening, F. (2011). Judgment and decision making in young children: Probability, expected value, belief updating, heuristics and biases. In M. K. Dhami, A. Schlottmann, & M. R. Waldmann (Eds.), *Judgement and decision making as a skill* (pp. 55–84). Cambridge, UK: Cambridge University Press.
- Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: Evidence for developmental trends. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 391–409). New York: Psychology Press.
- Schwarz, S., & Roebers, C. M. (2006). Age differences in the effects of social influence on children's eyewitness performance and their metacognitive monitoring. *Journal of Experimental Child Psychology*, 94, 229–248.
- Selmeczy, D., & Dobbins, I. G. (2013). Metacognitive awareness and adaptive recognition biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 678–690.
- Selmeczy, D., & Dobbins, I. G. (2017). Ignoring memory hints: The stubborn influence of environmental cues on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 1448–1469.
- Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., ... Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, 22, 264–271.