A Closer Look at Children's Metacognitive Skills: The Case of the Distinctiveness Heuristic

Introduction

The finding that distinctively encoded information – operationalized as the processing of differences relative to some context (Howe, 2006) – produces a memory advantage has a long and fruitful history in the domain of memory studies (see Hunt, 2013, for an overview). For instance, much research has shown that distinctive encoding usually improves correct memory and reduces false memory as compared to less distinctive encoding. This pattern has been demonstrated with various sorts of recognition tasks using, for example, picture vs. word stimuli (e.g., Gallo, Bell, Beier, & Schacter, 2006; Ghetti, Qin, & Goodman, 2002; Schacter, Israel, & Racine, 1999), reading words aloud vs. silently (e.g., Fawcett, Quinlan, & Taylor, 2012; Huff, Bodner, & Fawcett, 2015; Ozubko & MacLeod, 2010), salient vs. ordinary items (Strack & Bless, 1994), or bizarre vs. common events (e.g., Black et al., 2012).

Generally, studies conducted on adults suggest that at least two non-mutually exclusive editing processes – defined as mechanisms people use to avoid false memories (Lampinen & Odegard, 2006) – may account for the superior recognition performance in response to distinctive (e.g., pictures) over non-distinctive (e.g., words) material. First, memory decisions may be more accurate because distinctive encoding strengthens the memory trace, and thus produces high-quality memories. As a result, participants may be more likely to reject false information because they are able to recollect information that logically disconfirms the prior presentation of an item (e.g., "I couldn't have studied *pineapple,* because I remember studying *banana, raspberry,* and *peach,* and there were only three words per category"). This retrieval process is generally called "recall-to-reject" (Gallo, 2004; Gallo et al., 2006; Lampinen & Odegard, 2006; Rotello, Macmillan, & Van Tassel, 2000).

Second, the memory advantage may result from the implementation of the *distinctiveness heuristic*, a retrieval decision rule based on participants' metacognitive expectations about the quality of their memories (Dodson & Schacter, 2001; Schacter et al., 1999). According to this account, participants evaluate their memories against a criterion based on how detailed they expect those memories to be.

Specifically, people usually expect to be able to recollect more vivid details after distinctive encoding than after less distinctive encoding. When these expectations are not fulfilled, participants tend to conservatively decide that they have never encountered the stimulus before (reasoning that can be formalized as “if I had seen it, I would have remembered it”). Conversely, when participants do not have such metacognitive expectations – for example, after encoding a word – they are inclined to use a more liberal response criterion. Providing evidence for the critical role of expectations in participants’ memory decisions, Dodson and Schacter (2002) showed that the distinctiveness heuristic is used even if none of the items are distinctively encoded, as long as participants *believe* that some of the items were encoded in a distinctive manner (see also McDonough & Gallo, 2012). In sum, while the recall-to-reject process depends on the objective recollection of specific episodes, the distinctiveness heuristic mainly relies on subjective expectations about the quality of memory traces.

Interestingly, these two memory editing processes have been incorporated into various theoretical models of memory, such as fuzzy trace theory, the source-monitoring framework, and signal detection theory. According to fuzzy trace theory, people encode multiple representations of an item in parallel and these representations vary in terms of their precision. Memory traces that encode an item’s features are called “verbatim traces,” while traces that encode general meanings are called “gist traces” (Reyna & Brainerd, 1995). According to this theory, the recall-to-reject strategy would result from the retrieval of verbatim traces. Conversely, the distinctiveness heuristic would depend on awareness of the quality of the traces (Gomes & Brainerd, 2013).

Another theory that distinguishes between the recall-to-reject strategy and the distinctiveness heuristic is the source-monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993). According to this model, people accept a piece of information as having been previously presented only if they can attribute the source of this information to their memory, which is possible because memories for different sources contain characteristically different kinds of information. Within this framework, the recall-to-reject strategy is supposed to influence source-monitoring when people correctly attribute episodic details to their respective sources, leading to the rejection of misinformation. Conversely, the distinctiveness heuristic is assumed to influence source-monitoring when sources differ in the richness of their details, leading participants to generate different retrieval expectations.

Finally, the recall-to-reject strategy and the distinctiveness heuristic have also been interpreted within signal detection theory (Macmillan & Creelman, 2005). This framework characterizes how perceivers separate meaningful information (e.g., studied items) from noise (e.g., distractors). Because recall-to-reject is based on the ability to recollect logically inconsistent information, it is supposed to cause participants to better discriminate between studied items and distractors, increasing their sensitivity scores (i.e., a greater probability of responding "yes" when a studied item is presented and "no" when a new item is presented). On the other hand, given that the distinctiveness heuristic is based on metacognitive expectations, it is claimed to induce participants to set a stricter response criterion for distinctive than for less distinctive classes of stimuli (i.e., a greater tendency to respond "no" to an item whether it has been studied or not).

From a developmental perspective, how the two retrieval strategies, and particularly the distinctiveness heuristic, influence children's memory decisions is far from clear. Yet the distinctiveness effect could have a major influence on children's daily functioning, for instance, by helping them to distinguish fiction from real life. In the lab, a memory gain has been shown following the encoding of various distinctive materials in children aged from 4 to 10 years old (Geurten, Willems, & Meulemans, 2015a; Ghetti et al., 2002; Howe, 2006; Howe, Courage, Vernescu, & Hunt, 2000). To date, however, the processes underlying these distinctiveness effects have not been tested much in young children. Indeed, the distinctiveness heuristic has only recently been regarded as a potential explanation of the distinctiveness effect in children. According to the literature, metacognitive heuristics result from the automatization of the connection between a detected cue (e.g., metacognitive monitoring of the discrepancy between expected and actual quality of memory) and a behavioral decision (e.g., metacognitive control of memory decisions) (Geurten, Willems, & Meulemans, 2015b). Until recently, young children's metacognitive skills were believed to be too limited to implement such rules (Fritz, Howie, & Kleitman, 2010; Ghetti et al., 2002). Consistent with this view, a series of studies examining whether and how children rely on the distinctiveness heuristic (also known as the memorability-based heuristic) to reject false memories have found that 9-year-olds and adults more frequently rejected high- than low-memorability false events, but that 5- and 7-year-olds did not (Ghetti & Alexander, 2004; Ghetti & Castelli, 2006). However, other studies have now established that even children as young as 3 have basic monitoring and control abilities (e.g., Gerken, Balcomb, & Minton, 2011; Hembacher & Ghetti, 2014), paving the way for the possibility that young children may use certain monitored cues when making memory decisions.

To test this possibility, Geurten, Willems, and Meulemans (2015a) asked participants aged 4, 6, and 9 years old to study two lists of unrelated items. These two lists were presented either visually (pictures) or orally (words). After each study phase, participants performed recognition tests. The results revealed that even 4-year-olds produced fewer false recognitions in the picture condition than the word condition. Nevertheless, signal detection analyses indicated that the lower rate of false recognitions after picture encoding could potentially result from both the implementation of a conservative response criterion and an increase in children's ability to discriminate between studied items and lures. This pattern makes it difficult to decide whether metacognitive processes (creation of retrieval expectations leading to the use of the distinctiveness heuristic), memory processes (enhancement of memory traces allowing for the use of a recall-to-reject strategy), or both influenced children's recognition performance in the above-described study, insofar as each of these processes is anticipated to reduce participants' false recognition rate. In two recent studies, Moore and colleagues (Moore, Lampinen, Gallo, Adams, & Bridges, 2017; Moore, Lampinen, Gallo, & Bridges, 2017) examined the developmental trajectory of both the distinctiveness heuristic and the recall-to-reject strategy in 7- and 9-year-old children. They found that these two retrieval strategies improved differently with age, suggesting that their influence on the rejection of misinformation can potentially be differentiated.

According to Schacter et al. (1999; see also Forrin, Groot, & MacLeod, 2016), to fully establish that participants are relying on the distinctiveness heuristic to guide their memory decisions, an experimental paradigm manipulating the type of encoding not only between lists but also within the same list should be employed. Ordinarily, the use of the distinctiveness heuristic is somewhat ineffective following a mixed study list because such a design prevents participants from forming global metacognitive expectations about the kind of information they should be able to recollect. Indeed, in a mixed-list design, the absence of vivid memories does not guarantee that an item is new; it could simply have been presented in a less distinctive form during the encoding phase, preventing participants from tightening up their response criterion ("if I had seen it, I would have remembered it, but perhaps I just heard it"). For these reasons, if the conservative bias that has been observed in children with a pure-list design results from their reliance on the metacognitive rule, then a mixed-list manipulation should eliminate it. By contrast, increased discrimination due to an improved memory trace – for pictures as opposed to words – should remain unchanged. The fact that some items have been encoded in picture form and others in word form should not prevent children from recollecting logically inconsistent information more often after picture encoding than after word encoding. If such results are found, they would be very important for our understanding of both metacognitive and memory functioning in young children. Indeed, the finding that even young children can rely on the distinctiveness rule to regulate their memory decisions would provide interesting information on how metacognitive factors influence memory performance throughout childhood.

Overview of the Present Study

The primary aim of this study was to document the developmental course of distinctiveness effects throughout childhood. Specifically, we examined whether the reduction in the false recognition rate that is traditionally observed in children after distinctive encoding can be explained not only by enhanced discrimination between studied and new items but also by the implementation of a conservative response criterion resulting from the use of the distinctiveness heuristic.

To test this hypothesis, we conducted two experiments in which children aged 4–5, 6–7, and 8–9 years old were asked to study a set of items presented either in pictorial (distinctive) or in word (less distinctive) form. These age groups were selected because some recent studies have indicated that 4-year-olds already use certain metacognitive heuristics, while other studies have not found any reliance on metacognitive rules before the age of 8–9 years (see Geurten & Willems, 2016, for an overview). These inconsistencies make it necessary to further investigate whether and how children of those ages use the distinctiveness heuristic. In Experiment 1, pictures and words were displayed in two separate lists. In Experiment 2, both types of stimuli were presented within the same list.

Based on the results of Geurten, Willems, and Meulemans (2015a), we expected children in all age groups to be more discriminating and more conservative in the picture than in the word condition in Experiment 1. However, one of the main limitations on that study was that the encoding mode was confounded with the retrieval mode; namely, items encoded in pictorial form were also presented in pictorial form at test, making the picture condition much easier than the word condition. For this reason, in this study, test stimuli were presented in word form after both picture and word encoding.

In Experiment 2, several scenarios were possible. First, the conservative response bias observed in pure-list designs may be due not to the implementation of the distinctiveness heuristic but to the fact that lures produce lower levels of memory strength after distinctive than after less distinctive encoding. In that case, the more conservative bias observed for pictures than for words in the pure-list design should also be found in the mixed-list design. Second, there may be a developmental trend in the use of the distinctiveness heuristic, which accounts only for older participants' pattern of responses. In that case, the mixed-list manipulation should eliminate the more conservative response criterion observed for pictures as compared to words, but only in older children. Such findings would be consistent with previous studies (e.g., Ghetti & Alexander, 2004; Ghetti & Castelli, 2006) showing that children are not able to rely on the distinctiveness heuristic to regulate their memory performance before the age of 8–9. Third, the distinctiveness heuristic may account for the conservative response bias observed in pure-list designs at all ages. In that case, no differences should be found between the response bias scores for pictures and for words in any of our age groups. The latter pattern would provide evidence that children truly relied on their metacognitive expectations about distinctiveness to guide their memory decisions when the type of encoding was manipulated between lists. Whichever hypothesis is confirmed, the mixed-list design should not affect the quality of the memory trace, and consequently, children's ability to rely on the recall-to-reject strategy. For this reason, all age groups were anticipated to be more discriminating after picture than after word encoding.

Experiment 1

The aim of Experiment 1 was to replicate the results of Geurten, Willems, and Meulemans (2015a) by examining whether 4–5, 6–7, and 8–9-year-old children showed reduced false recognition rates after distinctive encoding than after less distinctive encoding. To this end, participants were asked to study two separate lists of items presented in either pictorial (picture condition) or word form (word condition), then to complete recognition tests. During each of the recognition tests, studied items were presented together with new items from three main categories (i.e., semantically, visually, or auditorily related to studied items; for another study using perceptually similar items as lures, see Lyle & Johnson, 2006).These categories of lures were selected because (a) they should increase the number of "yes" responses for new items and thus prevent ceiling effects, and (b) they should allow us to determine whether the amount of physical details shared by the studied items and the lures could influence children's memory decisions more after distinctive encoding than after less distinctive encoding. Specifically, if children use their expectations about the amount of physical details they should be able to recollect as a heuristic to guide their decisions, perceptually similar lures should lead to a higher rate of false recognition than semantically similar lures and unrelated lures, especially after distinctive encoding. Regardless of lure type, we expected participants both to discriminate better and to respond more conservatively in the picture condition than in the word condition.

Method

Participants. The participants were 42 typically developing children aged 4–5 (*M* = 59.99 months, *SD* = 4.36, 7 females), 6–7 (*M* = 80.03 months, *SD* = 7.51, 6 females) and 8–9 years (*M* = 107.15 months, *SD* = 6.60, 8 females). There were 14 children per group. Of these participants, 38 children were Caucasian and 4 children were originally from North Africa. The native language of all children was French and they were all from homes with middle- to upper-class socioeconomic status. They were recruited from preschools and elementary schools in the Province of Liège, Belgium. Data collection stopped when the number of participants was sufficient to reach a predicted power of .80 for a within-between interaction of medium effect size. This effect size was determined on the basis of previous studies conducted in the field (e.g., Geurten, Willems, & Meulemans, 2015a [?]; Ghetti & Alexander, 2004). All children were enrolled following written informed consent from their parents.

Materials. The stimuli consisted of two lists of 42 two-dimensional colored line drawings and their corresponding names (in French). Each list contained 18 randomly assigned study items and 24 lures. Following a procedure inspired by Geurten, Willems, and Meulemans (2015a), lures were divided into three categories: (a) 8 items that each looked like one specific studied picture but were unrelated to any of the studied words (visual lures/unrelated when presented in the word condition, e.g., Moon–Banana); (b) 8 items that each sounded like one specific studied word but were unrelated to any of the studied pictures (auditory lures/unrelated when presented in the picture condition, e.g., Moon–Spoon); and (c) 8 items that each had strong semantic links with one specific studied item (semantic lures, e.g., Moon–Sun). All stimuli were common objects or animals and were selected to be included in the vocabulary of 4-year-olds. More examples of studied items and lures can be found in Appendix A.

In all, 64 pictures were taken from the standardized data sets developed by Rossion and Pourtois (2004). The remaining 20 pictures were retrieved from a free internet database; half were assigned to each list either as studied items or as lures. On average, word stimuli were equal in terms of frequency (a mean of 15.6 occurrences per million words; Radeau, Mousty, & Content, 1990) and number of syllables (a mean of 2.01 syllables). Pictures contained similar amounts of detail.

Procedure. Institutional Review Board approval was obtained from the local ethics committee before data collection began (protocol number: 1516-21). Children were tested individually in a quiet room in their school. Each participant underwent a session lasting approximately 45 minutes. One of the two lists (list A or B) was randomly assigned to each of the two encoding modes (picture or word). List assignment to a type of item was counterbalanced. Specifically, half of the participants encoded list A as pictures and list B as words. The other half encoded list A as words and list B as pictures. Moreover, the order of lists was also counterbalanced between participants, with half of them encoding list A first and the other half encoding list B first. For each list, children went through two successive phases: (a) a study phase, and (b) a recognition phase. These two phases were separated by a 10-minute period that was filled with nonverbal cognitive tasks including the Matrix Reasoning test, a subtest of the Wechsler Intelligence Scale for Children (WISC-IV; Wechsler, 2005) and of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III; Wechsler, 2004) that was used to evaluate children's nonverbal intelligence, and the go/no-go inhibition test (Raaijmakers et al. 2008). For each participant, one of these two tests was randomly assigned to one of the two lists of items. There was a 2-minute delay between the study phase and the recognition phase for one item type (e.g., picture) and the study phase and the recognition phase for the other type (e.g., word). This period was filled with some conversation between the child and the experimenter.

Study phases. Each child studied a list of 18 stimuli (either pictures or words) in random order. Participants were instructed to try to remember as many items as possible so they could recall them later. During word encoding, participants were asked to look at a stimulus in the center of the screen ("+") while the study items were named by a recorded female voice at a rate of one every 2.5 s. During picture encoding, each stimulus appeared in the center of the screen for 2.5 s.

Recognition phases. Participants were told that they would be presented with both studied and new items. After word encoding, children were instructed to respond "yes" if they remembered hearing the stimulus during the study phase and "no" if they did not. After picture encoding, they were instructed to answer "yes" if they remembered seeing a picture of the item during the study phase and "no" otherwise. Test stimuli were presented in word form after both picture and word encoding (for a similar procedure, see Ghetti et al., 2002). Specifically, a crosshair appeared in the center of the screen while items included in the list were successively named in random order. After each response, a blank screen was presented for 500 ms, followed by the presentation of the next item. In each experimental condition, 18 studied items and 24 lures from three different categories were presented.

Measures. In both Experiments 1 and 2, the main measures were (a) hit rates for the two types of items (i.e., rates of correctly recognized studied items divided by the total number of studied items); (b) false recognition rates for the two types of items (i.e., rates of falsely recognized lures divided by the total number of lures); and (c) sensitivity (d') and response bias (C) scores, which were estimated by comparing the number of correctly identified studied items with the number of falsely recognized lures (Macmillan & Creelman, 2005).

Results

First, we examined differences in children's hit and false recognition rates after picture encoding and after word encoding. Then, we tested their sensitivity and response bias after these two types of encoding. Unless otherwise noted, differences were significant to at least $p < .05$ in both Experiments 1 and 2. However, as orthodox statistics did not allow us to distinguish between insensitive data (i.e., no significant results) and evidence of the null hypothesis (i.e., a real absence of differences between two conditions), we chose to use Bayesian statistics in addition to more classical frequentist analyses (Dienes, 2014). Bayes factors (*B*) indicate the relative strength of evidence for two theories (e.g., Jarosz & Wiley, 2014). These factors allow three types of conclusions: (a) there is strong evidence for the alternative hypothesis (*B* much greater than 1); (b) there is strong evidence for the null hypothesis (*B* close to 0); and (c) the evidence is insensitive (*B* close to 1). Lee and Wagenmakers (2014) proposed the following decision criteria: Bayes factors greater than 3 or less than 1/3 represent substantial evidence against or for the null hypothesis. Anything between 1/3 and 3 indicates that more evidence is needed. In the present experiments, Bayesian tests were conducted using the default conservative priors from our software program (JASP, 2014; see van de Schoot & Depaoli, 2014).

Preliminary analyses indicated that no gender or order effects were significant for any of the dependent variables. No group differences were found regarding parental education and children's nonverbal intelligence (*Fs* < 1.7, respectively assessed using the two parents' years of education and scores on the Matrix Reasoning test; Wechsler, 2004, 2005).

Hit and false recognition rates.

Hit rate. Hit rate was analyzed with a mixed-design ANOVA, 3 (Age: 4–5, 6–7, 8–9 years old) x 2 (Item Type: picture, word), with Item Type as within-participants factor. No main or interaction effects reached significance, all *Fs* < 2.6, *ps* > .09. Bayesian analyses of the hit rate provided evidence in favor of the null hypothesis for each effect, all *Bs* < .31.

False recognition rate. False recognition rate was analyzed with a mixed-design ANOVA, 3 (Age: 4–5, 6–7, 8–9 years old) x 2 (Item Type: picture, word) x 3 (Lure Type: visual/auditory, semantic, unrelated), with age as the only between-participants factor. Data are presented in Table 1. No significant differences were found between our three age groups, $F = 2.41$, $p = .10$. However, the results indicated that participants had a higher false recognition rate for the word items ($M = .31$) than for the picture items ($M = .16$), $F(1,39) = 10.95$, $MSe = 5.04$, $\eta^2_p = .22$. Children's false recognition rates also depended on the type of lure, $F(2,78) = 6.88$, $MSe = 0.83$, $\eta^2_p = .18$. Planned comparisons indicated that participants had a lower false recognition rate for unrelated lures ($M = .20$) than for semantic lures ($M = .24$), $F(1,39) = 6.74$, $MSe = 0.55$, $p = .013$, or visual/auditory lures ($M = .25$), $F(1,39) = 26.56$, $MSe = 0.25$, $p < .001$. No differences were found between semantic and visual/auditory lures, $F < 1$. Finally, an Item Type x Lure Type interaction also proved significant, $F(2,78) = 5.25$, $MSe = 0.79$, $p = .012$, indicating that children had a higher false recognition rate after word encoding than after picture encoding for both semantic ($M = .32$ and .15, for words and pictures, respectively), $F(1,39) = 11.11$, $MSe = 2.20$, $p = .002$, and unrelated lures ($M = .29$ and .10), $F(1,39) = 17.53$, $MSe = 1.59$, $p < .001$, but not for visual/auditory lures ($M = .30$ and .21), $F = 3.42$. The latter finding is particularly interesting because it indicates that the type of encoding affects how children reject lures that share physical features with studied items. No other interaction effects were found, all $Fs < 1.25$, $ps > .30$. Bayesian statistics indicated that more evidence is needed before concluding that there is no difference between age groups ($B = .61$) but revealed evidence in favor of an effect of Item Type ($B = 5.61$) and of Lure Type ($B = 3.26$), and an Item Type x Lure Type interaction ($B = 3.09$). Evidence in favor of the null hypothesis was found for the other effects (all $Bs < .23$).

Table 1. Proportions of Correctly Recognized Studied Items (Hits) and Falsely Recognized Lures (False Recognitions), and Values for Sensitivity (d') and Response Bias (C) by Age Group for the Two Item Types in Experiment 1

	4–5 years (n = 14)		6–7 years (n = 14)		8–9 years (n = 14)		All (n = 42)	
	Picture	Word	Picture	Word	Picture	Word	Picture	Word
Hits	0.54	0.48	0.44	0.54	0.44	0.63	0.48	0.55
	(.27)	(.27)	(.25)	(.23)	(.25)	(.16)	(.26)	(.23)
False recognitions								
Total	0.27	0.35	0.13	0.33	0.07	0.23	0.16	0.31
	(.28)	(.31)	(.11)	(.29)	(.09)	(.27)	(.19)	(.29)
Visual/Auditory	0.33	0.39	0.18	0.30	0.13	0.21	0.21	0.30
	(.31)	(.32)	(.16)	(.28)	(.13)	(.22)	(.23)	(.28)
Semantic	0.27	0.33	0.14	0.38	0.05	0.27	0.15	0.32
	(.29)	(.33)	(.15)	(.31)	(.09)	(.30)	(.21)	(.31)
Unrelated	0.21	0.32	0.05	0.32	0.03	0.21	0.10	0.29
	(.26)	(.34)	(.06)	(.32)	(.07)	(.29)	(.18)	(.31)
Sensitivity (d')	1.02	0.43	1.14	0.69	1.51	1.29	1.22	0.80
	(.72)	(.81)	(.68)	(.85)	(1.04)	(1.03)	(.83)	(.89)
Response bias (C)	0.26	0.22	0.72	0.22	0.85	0.30	0.60	0.25
	(.88)	(.94)	(.56)	(.71)	(.43)	(.41)	(.69)	(.75)

Note. *Standard deviations are in parentheses. Visual/Auditory = when children encoded picture items, they were presented with the visual (look-alike) lures at test. When children encoded word items, they were presented with the auditory (sound-alike) lures at test.*

Signal detection analyses. We performed a signal detection analysis to determine the contributions of sensitivity and response bias to children's reduced false recognition rate with pictures compared to words (Macmillan & Creelman, 2005). The sensitivity (*d'*) and response bias (*C*) scores were analyzed with mixed-design ANOVAs, 3 (Age) x 2 (Item Type), with Item Type as the within-participants factor. The results are presented in Table 1.

Sensitivity. Children's ability to discriminate between studied items and lures improved with age, $F(2,39) = 4.35$, $MSe = 30.80$, $\eta^2_p = .18$. Planned comparisons revealed that 9-year-olds ($M = 1.40$) were better at discriminating than 4-year-olds ($M = .72$), $F(1,39) = 8.20$, $MSe = 30.79$, $p = .007$, or 6-year-olds ($M = .92$), $F(1,39) = 4.20$, $MSe = 30.80$, $p = .047$. No differences were found between the two younger groups, $F < 1$. The results also indicated that participants discriminated between studied items and lures better after picture encoding ($M = 1.22$) than after word encoding ($M = .80$), $F(1,39) = 6.21$, $MSe = 23.02$, $\eta^2_p = .14$. No interaction effects were found, $F = 0.40$, $p = .67$. Bayesian analyses provided evidence in favor of the alternative hypothesis for both the effect of age ($B = 3.85$) and the effect of Item Type ($B = 3.452$). Evidence in favor of the null hypothesis was found for the interaction effect ($B = .33$).

Response bias. Participants responded more conservatively for the picture items ($M = .60$) than for the word items ($M = .25$), $F(1,39) = 6.01$, $MSe = 16.43$, $\eta^2_p = .13$ (see Figure 1, Panel A). No other effects reached significance, all *Fs* < 1.54, *ps* > .23. Bayesian analyses produced evidence of a difference between the response bias scores for pictures and for words ($B = 3.857$) and evidence in favor of the null hypothesis for the age effect ($B = .34$). More evidence is needed for the Age x Item Type interaction ($B = .649$).

Discussion

Experiment 1 provides important information about processes that may be involved in children's recognition decisions after distinctive encoding. Specifically, we replicated the results of an earlier study (Geurten, Willems, & Meulemans, 2015a) by showing a decrease in children's false recognition rate for picture items compared with word items. This was done by improving the method used in the previous study. Indeed, in the present experiment, the encoding mode was not confounded with the retrieval mode. The signal detection analysis that we carried out indicated that this reduction could be explained both by an increase in children's ability to discriminate between studied items and lures – which is consistent with the hypothesis that the memory trace is enhanced, allowing for the use of a recall-to-reject strategy – and by the implementation of a conservative response criterion – which is consistent with the hypothesis that a metacognitive heuristic is used.

Interestingly, the finding that the false recognition rate was significantly lower after picture encoding than after word encoding for semantic and unrelated lures, but that no difference was found for visual/auditory lures, is also coherent with the metacognitive hypothesis. The visual proximity of studied pictures and visual lures may have induced participants not to reject these lures simply because the amount of physical details they recollected about these visually similar lures was sufficient to satisfy their metacognitive expectations, resulting in an equal rate of false recognitions for pictures and for words. Although relatively speculative, this hypothesis is consistent with previous work indicating that vivid episodic details are sometimes associated with false recognition when people borrow or import features (e.g., color, shape) from one item into the memory for another item (Lampinen, Meier, Arnal, & Leding, 2005; Lyle & Johnson, 2006). Other investigations are, of course, needed to further test this hypothesis in children.

To a lesser degree, this reasoning may also apply to semantic lures, which may explain why 4-year-old children showed a numerically smaller difference between response bias scores for pictures and for words than older children. Indeed, a classical finding in the literature is that young children's recognition decisions are influenced very little by semantic associations between items (e.g., Brainerd, Reyna, & Ceci, 2008). Still, items that are semantically related also frequently share physical features (e.g., Ladle – Spoon), possibly inducing young participants – for whom the semantic association between items was not at the forefront – to accept more semantic lures on the basis of the amount of physical details they recollected. To test this, we recomputed the sensitivity and response bias scores using only false recognitions for unrelated lures. Overall, our finding remained essentially unchanged. However, the difference in young children's response bias scores for pictures and for words (M_{Diff} = .16) appeared to be much larger than when all three types of lures were taken into account (M_{Diff} = .04), and to be statistically similar to that observed in 6-year-old (M_{Diff} = .24) and 9-year-old (M_{Diff} = .26) children. Sensitivity and response bias scores for each type of lure are provided as supplemental results.

Nonetheless, our results are not sufficient in themselves to conclude that participants relied on the distinctiveness heuristic when making memory decisions. To further explore whether the distinctiveness heuristic is involved in children's recognition decisions, we conducted a second experiment in which we put participants in a position where they could not use their metacognitive expectations about the kinds of information they felt they should be able to recollect to guide their memory judgments.

Experiment 2

Our goal in Experiment 2 was to determine whether the reduced false recognition rate observed after picture encoding in Experiment 1 resulted from the implementation of the distinctiveness heuristic or was due only to an increase in children's ability to discriminate between studied items and lures. To test this, we decided to use a procedure that rules out the use of metacognitive expectations, and thus the implementation of the distinctiveness rule. Specifically, we manipulated picture and word encoding on a within-list basis. A mixed-list design is generally expected to prevent the use of the distinctiveness heuristic because the lack of recollection of physical details does not necessarily indicate that an item is a lure but may suggest that it was studied in a less distinctive form (Schacter et al., 1999). In this context, if the reduction in false recognitions obtained in Experiment 1 was due to the use of the metacognitive heuristic, we would expect differences in response bias between picture stimuli and word stimuli to be eliminated in Experiment 2.

Method

Participants. The participants were 42 typically developing children aged 4–5 (*M* = 57.03 months, *SD* = 6.41), 6–7 (*M* = 86.49 months, *SD* = 8.42) and 8–9 years (*M* = 111.90 months, *SD* = 4.12). There were 14 children per group (7 girls and 7 boys). All of these participants were Caucasian and from homes with middle- to upper-class socioeconomic status. Their native language was French. All participants were recruited from preschools and elementary schools in the Province of Liège, Belgium.

Materials and procedure. The study and test stimuli were identical to those used in Experiment 1, but the manner in which they were presented differed. In Experiment 2, both classes of stimuli (pictures and words) were mixed within the same list. During the study phase, children studied 36 items. Half of the items were presented in pictorial form and the other half in word form. Across participants, each item appeared equally often as a picture and as a word. In Experiment 1, when confronted with a pure list, participants could have adopted an encoding approach that relied either on visual (picture items) or on phonological (word items) characteristics of the stimuli. This could have boosted their memory. In Experiment 2, our aim was to reduce the children's ability to form metacognitive expectations, but not to reduce the quality of their memory, by preventing the use of this encoding strategy. For this reason, we chose to group items into four series of nine stimuli. Two series included randomly selected pictures and two series included randomly selected words. For each participant, two series never appeared consecutively in the same presentation mode (see Schacter et al., 1999, for a similar procedure). Consequently, participants still had the opportunity to encode items on the basis of their perceptual characteristics. During the study phase, all 84 items (36 studied items and 48 lures) in the recognition phase were presented as auditory words. There were 24 lures from three categories that were related to the picture stimuli and 24 lures from three categories that were related to the word stimuli.

As each lure was associated with either a specific picture or a specific word, we were able to obtain separate mixed-picture and mixed-word false recognition rates. In studies that used a mixed-list design without associating new items with specific studied items, a single false recognition rate had to be computed for both pictures and words because there was no way of determining whether a given false recognition occurred because a new item was mistaken for a picture stimulus or for a word stimulus (Forrin et al., 2016). The procedure used in this experiment – involving visual, auditory, or semantic associations between a specific lure and a specific studied item – allowed us to avoid this issue without using the Deese-Roediger-McDermott (DRM) paradigm, which is not very appropriate for studying young children's false memories (e.g., Metzger et al., 2008).

Results

We conducted the same analyses as in Experiment 1. Mean hit rate, mean false recognition rate, *d'* scores, and *C* scores are presented in Table 2. Preliminary analyses indicated no gender or order effects on any of the dependent variables. No group differences were found regarding parental education and nonverbal intelligence, $Fs < 1.4$.

Hit and false recognition rates.

Hit rate. Hit rate was analyzed with mixed-design ANOVAs, 3 (Age) x 2 (Item Type), with Item Type as the within-participants factor. Participants had a higher hit rate for pictures ($M = .45$) than for words ($M = .40$), $F(1,39) = 5.10$, $MSe = 0.47$, $\eta^2_p = .12$. No other main or interaction effects were significant, all $Fs < 1.27$, $ps > .29$. Bayesian statistics produced evidence in favor of the null hypothesis for the effect of age ($B = .31$) and the Age x Item Type interaction ($B = .34$). Our results also indicated that more evidence is needed regarding the Item Type effect ($B = 1.48$).

False recognition rate. False recognition rate was analyzed with mixed-design ANOVAs, 3 (Age) x 2 (Item Type) x 3 (Lure Type), with age as the only between-participants variable. In contrast with Experiment 1, no differences were found between the rate of false recognitions for pictures and for words, $F = 1.65$, $p = .21$. However, the false recognition rate decreased with age, $F(2,39) = 9.37$, $MSe = 3.73$, $\eta^2_p = .32$. Planned comparisons revealed that 4-year-olds ($M = .24$) had a higher rate of false recognitions than 6-year-olds ($M = .07$), $F(1,39) = 12.21$, $MSe = 3.73$, $p = .001$, or 9-year-olds ($M = .05$), $F(1,39) = 15.69$, $MSe = 3.72$, $p < .001$. No differences were found between the two older groups, $F < 1$. An Age x Lure Type interaction was also significant, $F(4,39) = 3.33$, $MSe = 0.30$, $\eta^2_p = .15$, indicating that 4-year-old children had a lower rate of false recognitions for semantic lures ($M = .21$) than for visual/auditory lures ($M = .25$), $F(1,39) = 6.65$, $MSe = 0.13$, $p = .014$, while 9-year-old children showed a trend toward a higher rate of false recognitions for semantic lures ($M = .07$) than for visual/auditory lures ($M = .04$), $F(1,39) = 4.02$, $MSe = 0.13$, $p = .052$, and a higher rate of false recognitions for semantic lures ($M = .07$) than for unrelated lures ($M = .03$), $F(1,39) = 5.96$, $MSe = 0.15$, $p = .019$. No differences were found between the types of lures for 6-year-old children, all $Fs < 2.05$, $ps > .13$. These results are consistent with classical findings in the literature on children's memory performance, suggesting that the influence of semantic associations between items on recognition performance increases with age (e.g., Brainerd et al., 2008). No other main or interaction effects reached significance, all $Fs < 1.64$, $ps > .35$. Bayesian analyses of the false recognition rate revealed evidence in favor of a difference between age groups ($B = 26.63$) and for the Age x Lure Type interaction ($B = 3.09$). Evidence in favor of the null hypothesis was found for the other main and interaction effects, all $Bs < .04$.

Table 2. Proportions of Correctly Recognized Studied Items (Hits) and Falsely Recognized Lures (False Recognitions), and Values for Sensitivity (d') and Response Bias (C) by Age Group for the Two Item Types in Experiment 2

	4–5 years (n = 14)		6–7 years (n = 14)		8–9 years (n = 14)		All (n = 42)	
	Picture	Word	Picture	Word	Picture	Word	Picture	Word
Hits	0.43	0.42	0.44	0.33	0.49	0.44	0.45	0.40
	(.14)	(.14)	(.17)	(.13)	(.20)	(.16)	(.17)	(.15)
False recognitions								
Total	0.22	0.26	0.06	0.08	0.05	0.04	0.11	0.13
	(.17)	(.22)	(.10)	(.14)	(.04)	(.06)	(.13)	(.18)
Visual/Auditory	0.25	0.26	0.05	0.07	0.04	0.04	0.12	0.12
	(.20)	(.24)	(.11)	(.14)	(.06)	(.06)	(.16)	(.19)
Semantic	0.17	0.26	0.09	0.08	0.08	0.06	0.11	0.13
	(.15)	(.22)	(.10)	(.14)	(.06)	(.08)	(.12)	(.18)
Unrelated	0.22	0.26	0.04	0.08	0.03	0.04	0.10	0.13
	(.18)	(.23)	(.11)	(.14)	(.05)	(.08)	(.15)	(.18)
Sensitivity (d')	0.78	0.58	1.49	1.15	1.69	1.59	1.32	1.11
	(.55)	(.60)	(.69)	(.71)	(.78)	(.28)	(.77)	(.69)
Response bias (C)	0.53	0.56	0.91	1.03	0.82	0.93	0.77	0.83
	(.50)	(.52)	(.30)	(.37)	(.31)	(.38)	(.40)	(.47)

Note. *Standard deviations are in parentheses. Visual/Auditory = when children encoded picture items, they were presented with the visual (look-alike) lures at test. When children encoded word items, they were presented with the auditory (sound-alike) lures at test.*

Signal detection analyses. As in Experiment 1, we analyzed *d'* and *C* scores with mixed-design ANOVAs, 3 (Age) x 2 (Condition), with condition as within-participants factor.

Sensitivity. Participants' ability to discriminate between studied items and lures increased substantially with age, $F(2,39) = 12.75$, $MSe = 20.45$, $\eta^2_p = .40$. Planned comparisons indicated that the 4-year-olds ($M = .68$) were less able to discriminate than the 6-year-olds ($M = 1.32$), $F(1,39) = 10.87$, $MSe = 20.45$, $p = .002$, or the 9-year-olds ($M = 1.64$), $F(1,39) = 24.61$, $MSe = 20.45$, $p < .001$. No differences were found between the two older groups, $F = 2.77$, $p = .09$. Furthermore, participants tended to have higher sensitivity scores for picture items ($M = 1.32$) than for word items ($M = 1.11$), $F(1,39) = 3.74$, $MSe = 9.80$, $p = .06$, $\eta^2_p = .09$. As was observed in Experiment 1, the Age x Item Type interaction was not significant, $F = .39$, $p = .68$. Bayesian analyses provided evidence in favor of an effect of age ($B = 319.41$) and Item Type ($B = 3.05$). Further evidence is needed regarding the Age x Item Type interaction ($B = .49$).

Response bias. In contrast with Experiment 1, there were no differences between response bias scores for pictures and for words, $F = 1.19$, $p = .31$, but the results indicated that response bias increased with age, $F(2,39) = 4.74$, $MSe = 11.23$, $\eta^2_p = .20$. Specifically, 4-year-old children ($M = .55$) were less conservative than 6-year-old children ($M = .97$), $F(1,39) = 8.60$, $MSe = 11.23$, $p = .006$, or 9-year-old children ($M = .88$), $F(1,39) = 5.20$, $MSe = 11.23$, $p = .028$. No differences were found between 6- and 9-year-old children, $F < 1$. No other effects were significant, $F = 1.75$; $p = .19$ (see Figure 1, Panel B). Despite their clarity, these results assessed only the likelihood of differences between age and item type. By using a Bayesian method, we can test the opposing hypothesis that the two types of items are equivalent (Dienes, 2014). Bayesian analyses of the response bias scores produced evidence in favor of the null hypothesis both for the Item Type ($B = .31$) and the Age x Item Type ($B = .32$) effects.

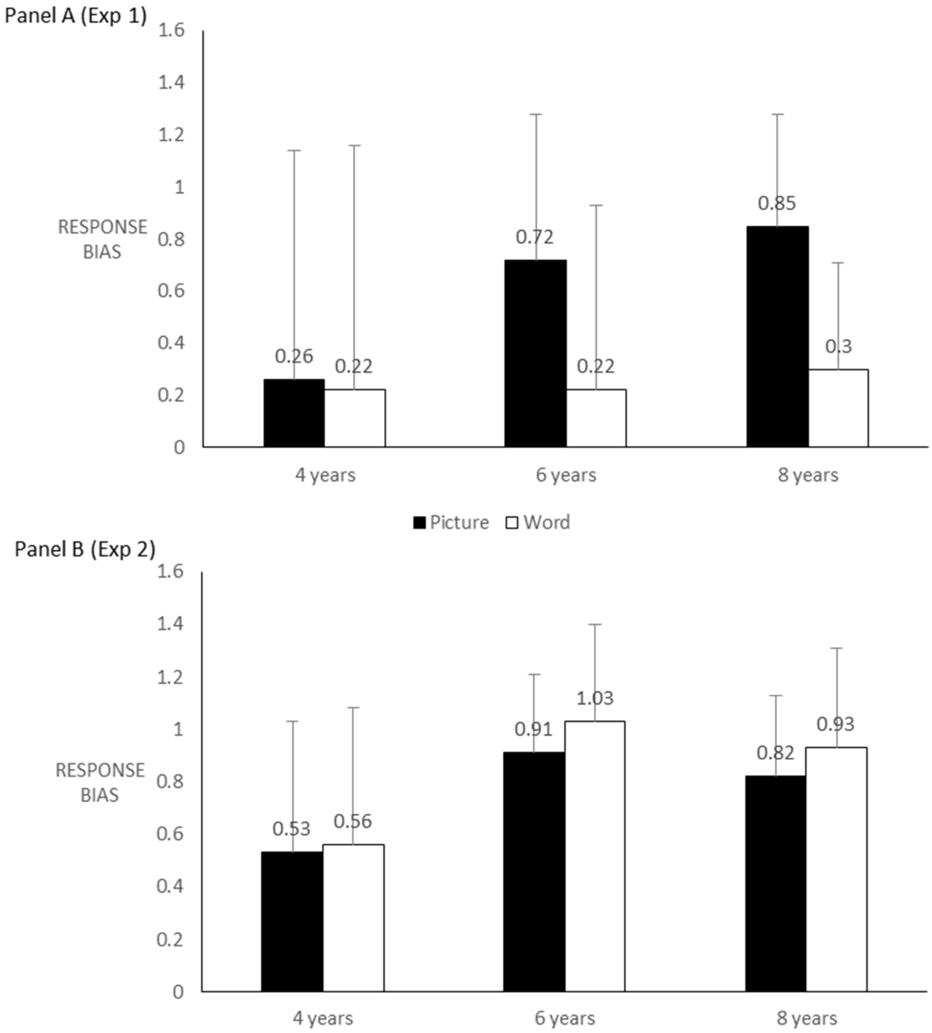


Figure 1. Response bias scores by age groups for the two types of items in both experiments. Error bars are standard deviations.

Discussion

The reduced rate of false recognitions observed after picture encoding in Experiment 1 was not found in Experiment 2. Indeed, no differences were revealed between the false recognition rates of lures associated with studied pictures and of lures associated with studied words. Similarly, the more conservative response bias observed for all age groups in the picture encoding condition than in the word encoding condition in Experiment 1 was not found in Experiment 2. Importantly, the Bayesian statistics indicate that this absence of effect was not due to a lack of power but provides evidence for a real lack of difference between picture and word items. According to Schacter et al. (1999), the fact that the reduced false recognition rate is eliminated when participants have access to list-specific information (i.e., tracking the memory source) before they are able to determine whether recollection of vivid details should be expected supports the idea that recognition decisions after picture encoding depend, at least partially, on people's metacognitive expectations.

General Discussion

A memory benefit is generally observed after the encoding of distinctive information as compared with less distinctive information. In children, however, especially young children, the processes underlying this distinctiveness effect have remained mostly unexplored. The present study contributed to this field by examining whether the reduction in the false recognition rate commonly observed after distinctive encoding results from an increase in children's ability to discriminate between studied items and lures or from the use of the distinctiveness heuristic.

Distinctiveness Heuristic

Replicating the results obtained by Geurten, Willems, and Meulemans (2015a), Experiment 1 indicated that even young children showed a lower false recognition rate in the picture than in the word condition. The signal detection analysis revealed that the lower false recognition rate after picture encoding was due to both an increase in sensitivity scores and the implementation of a conservative response criterion, which is consistent with the involvement of both a memory and a metacognitive process. To further explore which of these two processes accounts for the pattern of results observed in Experiment 1, we created conditions in which the use of the distinctiveness heuristic would not suppress false recognitions after picture encoding. Specifically, we presented participants with picture and word stimuli within the same list (Dodson & Schacter, 2001; Schacter et al., 1999). Consistent with the heuristic hypothesis, the results of Experiment 2 indicated no differences between false recognition rates and response bias scores for pictures and for words.

Overall, Experiments 1 and 2 seem to provide evidence that children as young as 4 rely on the distinctiveness heuristic to guide their memory decisions, resulting in a decrease in the false recognition rate when items are presented using a pure-list design (Experiment 1) but not when they are presented using a mixed-list design (Experiment 2). These findings may have important implications for our understanding of the development of metacognition and the involvement of metacognitive skills in children's memory performance.

Until recently, authors studying metacognition assumed that metacognitive skills emerged only relatively late in child development (Fritz et al., 2010). Consequently, metacognition was not expected to have a significant influence on children's memory performance before the age of 7 or 8 years (see Schneider & Lockl, 2008, for an overview). Recent results, however, indicate that children's ability to evaluate the current state of their memory operations (metacognitive monitoring) and to regulate these operations on the basis of this evaluation (metacognitive control) develops much earlier than had previously been thought, possibly as early as age 3 (Gerken et al., 2011; Hembacher & Ghetti, 2014; Lipowski, Merriman, & Dunlosky, 2013). Moreover, some emerging data show that 4-year-old children are already able to use certain monitored cues to guide their memory decisions, especially when those decisions are made automatically, on the basis of metacognitive heuristics (e.g., Geurten, Lloyd, & Willems, 2016; Geurten, Meulemans, & Willems, 2015; Geurten, Willems, Germain, & Meulemans, 2015). Our current findings, indicating that young children may use their expectations about the quality of their memories when making recognition decisions, add to these data and further confirm that basic metacognitive skills influence children's memory performance long before the middle years of childhood.

From a developmental perspective, our results suggest that all children rely on the distinctiveness heuristic in the same way when pictures and words are presented in separate lists. Indeed, no age or age x item type differences were found in children's false recognition rates or response bias scores in Experiment 1. However, a closer inspection of the means seems to indicate a very small difference in response bias scores for picture and word items in 4-year-old children. Further analyses revealed that this difference increased when only unrelated lures were used to calculate response bias scores, suggesting that young children may have more difficulties than older children in determining how much information they should be able to recollect after distinctive encoding. Indeed, in this study, lures have to share virtually no physical features with studied items (i.e., unrelated lures) for young children to reject them. These findings are important because they suggest that, even though the distinctiveness heuristic is applied from age 4 onward, its use still improves with age. Interestingly, this might explain why the distinctiveness effect had previously been shown to be smaller in young than in older children (e.g., Ghetti & Alexander, 2004; Howe et al., 2000; Moore, Lampinen, Gallo, Adams, et al., 2017; Moore, Lampinen, Gallo, & Bridges, 2017). Moreover, age-related differences were found in response bias scores in Experiment 2; specifically, 6- and 9-year-old children responded more conservatively than 4-year-old children to both pictures and words when items were presented within the same list. This pattern is consistent with findings presented elsewhere in the literature indicating that children naturally become more conservative with age (Geurten, Willems, & Meulemans, 2015a). Overall, these results seem to signal age-related changes in children's metacognitive expectations about the quality of their memory. Future studies should be conducted to determine whether these changes in metacognitive expectations result from memory development (i.e., because older children are able to remember more details than young children, they start to develop higher expectations), from metacognitive development (i.e., older children are able to make a finer-grained analysis of what kinds of episodic details they should be able to recollect after distinctive encoding), or both.

Distinctiveness Heuristic and Recall-to-Reject

Our intention here is not to claim that the use of the distinctiveness heuristic fully accounts for the memory advantage observed after picture encoding. Our data indicate that children are more discriminating after picture encoding than after word encoding even when the two types of items are mixed within the same list. According to the literature, this finding is consistent with the idea that distinctive encoding produces high-quality memories that make participants more likely to recall logically inconsistent information that they can then use to reject lures. This retrieval memory process – which is generally called "recall-to-reject" (e.g., Gallo et al., 2006; Rotello et al., 2000) – probably operates concurrently with metacognitive processes when items are presented in a pure-list design. However, while the recall-to-reject process appears to remain effective when items are presented in a mixed-list design, the distinctiveness heuristic does not. This dissociation between memory (i.e., "recall-to-reject" strategy) and metacognitive (i.e., distinctiveness heuristic) retrieval processes is interesting because it extends the findings of Moore et al. (Moore, Lampinen, Gallo, Adams, et al., 2017; Moore, Lampinen, Gallo, & Bridges, 2017) indicating that children's retrieval strategies based on *high-quality* memories or on *expectations* about high-quality memories are influenced differently by the context and the task's characteristics. In the same vein, the fact that children's ability to discriminate between studied items and lures increases with age, especially in Experiment 2 where the larger number of stimuli makes the recognition task more difficult, suggests that the use of the "recall-to-reject" strategy improves with age.

This study was designed and analyzed within the signal detection theory framework. However, from a theoretical perspective, our results are also consistent with other classical models such as fuzzy trace theory (Reyna & Brainerd, 1995) and the source-monitoring framework (Johnson et al., 1993), which postulate that related, but distinct, editing mechanisms help people to control the quality of their memory. For instance, the fact that picture encoding leads to overall better memory performance in both experiments is consistent with the hypothesis that distinctive encoding favors the retrieval of verbatim traces – which include the items' features – as postulated by fuzzy-trace theory. Similarly, the fact that the distinctiveness heuristic appears to be used only when a between-list design is employed seems consistent with the source-monitoring framework whereby being in a position to distinguish between sources with different amounts of details is an essential prerequisite to form different retrieval expectations, and thus to implement the metacognitive heuristic.

Limitations and Conclusion

One of the main limitations of this study is the small sample sizes of both experiments. Although these sample sizes were determined on the basis of a power analysis and Bayesian statistics were used to distinguish between insensitive data and evidence of the null hypothesis, the absence of an age effect on the hit and false recognition rates remains surprising. However, the fact that an age effect was found for the sensitivity score – which takes into account both hit and false recognition rates – in both experiments 1 and 2 suggests that there were age-related improvements in children's memory performance but that our design was not powerful enough to reveal them when hits and false recognitions were examined separately. Nonetheless, future studies should be conducted with larger samples to confirm this hypothesis and replicate our developmental findings.

In conclusion, the results of the two experiments conducted in this study indicate that, in children, the memory benefit traditionally enjoyed after distinctive encoding is not only due to an enhancement of the memory trace but also results from the implementation of a distinctiveness heuristic based on participants' metacognitive expectations. This suggests that, in some instances, even 4-year-olds can rely on basic metacognitive skills to heuristically reduce their rate of false recognitions. However, these findings need to be replicated and generalized to other types of distinctive material. As this research focuses on the study of differences in recognition performance between pictures and words, the distinctiveness effect is completely confounded with the encoding mode. It is therefore possible that our findings result not from a general distinctiveness effect but from a picture superiority effect that would not be generalized to other types of stimuli. Previous studies in adults have shown that similar patterns of results can be obtained by contrasting different classes of material (e.g., Dodson & Schacter, 2002), but this remains to be tested in children. Consequently, our results should not be generalized until they have been replicated using other classes of stimuli. Nevertheless, our data are important because they add to the small (but growing) amount of evidence indicating that metacognition develops very early in childhood and already has a positive influence on memory performance in the preschool years.

References

Black, S. R., McCown, S., Lookadoo, R. L., Leonard, R. C., Kelley, M., DeCoster, J., . . . Spence, S. A. (2012). Aging, imagery, and the bizarreness effect. *Aging, Neuropsychology, and Cognition, 19*, 566–591. doi:10.1080/13825585.2011.639868

Brainerd, C., Reyna, V., & Ceci, S. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin, 134*, 343–382.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781. doi:10.3389/fpsyg.2014.00781

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin and Review, 8*, 155–161. doi:10.3758/BF03196152

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language, 46*, 782–803.

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory, 20*, 655–666.

Forrin, N., Groot, B., & MacLeod, C. (2016). The d-prime directive: Assessing costs and benefits in recognition by dissociating mixed-list false alarm rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1090–1111. doi:10.1037/xlm0000214

Fritz, K., Howie, P., & Kleitman, S. (2010). "How do I remember when I got my dog?" The structure and development of children's metamemory. *Metacognition and Learning, 5*, 207–228. doi:10.1007/s11409-010-9058-0

Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 120–128. doi:10.1037/0278-7393.30.1.120

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory, 14*, 730–741.

Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid “labouring in vain” by attending more to learnable than unlearnable linguistic patterns. *Developmental Science, 14*, 972–979. doi:10.1111/j.1467-7687.2011.01046.x

Geurten, M., Lloyd, M.E., & Willems, S. (2016). Hearing “quack” and remembering a duck: Evidence for fluency attribution in young children. *Child Development, 88,* 514–522. doi:10.1111/cdev.12614

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

Geurten, M., & Willems, S. (2016). Metacognition in early childhood: Fertile ground to understand memory development? *Child Development Perspectives, 10,* 263–268.

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. (2015a). Are children conservative, liberal, or metacognitive? Preliminary evidence for the involvement of the distinctiveness heuristic in decision making. *Journal of Experimental Child Psychology, 132*, 230–239. doi:10.1016/j.jecp.2014.12.010

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

Ghetti, S., & Alexander, K. W. (2004). “If it happened, I would remember it”: Strategic use of event memorability in the rejection of false autobiographical events. *Child Development, 75*, 542–561. doi:10.1111/j.1467-8624.2004.00692.x

Ghetti, S., & Castelli, P. (2006). Developmental differences in false-event rejection: Effects of memorability-based warnings. *Memory, 14*, 762–776. doi:10.1080/09658210600648548

Ghetti, S., Qin, J., & Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective experience. *Developmental Psychology, 38*, 705–718. doi:10.1037/0012-1649.38.5.705

Gomes, C. F. A., & Brainerd, C. J. (2013). Dual processes in the development of reasoning: The memory side of the theory. In P. Barrouillet & C. Gauffroy (Eds.), *The development of thinking and reasoning* (pp. 221–242). New York: Psychology Press.

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. doi:10.1177/0956797614542273

Howe, M. L. (2006). Developmental invariance in distinctiveness effects in memory. *Developmental Psychology, 42*, 1193–1205. doi:10.1037/0012-1649.42.6.1193

Howe, M. L., Courage, M. L., Vernescu, R., & Hunt, M. (2000). Distinctiveness effects in children's long-term retention. *Developmental Psychology, 36*, 778–792. doi:10.1037/0012-1649.36.6.778

Huff, M. J., Bodner, G. E., & Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: A meta-analytic review of costs and benefits and their origins in the DRM paradigm. *Psychonomic Bulletin and Review, 22*, 349–365. doi:10.3758/s13423-014-0648-8

Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science, 22*, 10–15. doi:10.1177/0963721412463228

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, *7*, 2–9. doi:10.7771/1932-6246.1167

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28. doi:10.1037/0033-2909.114.1.3

Lampinen, J. M., Meier, C. R., Arnal, J. D., & Leding, J. K. (2005). Compelling untruths: Content borrowing and vivid false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 954–963. doi:10.1037/0278-7393.31.5.954

Lampinen, J. M., & Odegard, T. N. (2006). Memory editing mechanisms. *Memory, 14*, 649–654. doi:10.1080/09658210600648407

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.

Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology, 49*, 1505–1516. doi:10.1037/a0030614

Lyle, K., & Johnson, M. (2006). Importing perceived features into false memories. *Memory, 14*, 197–213. doi:10.1080/09658210544000060

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.

McDonough, I. M., & Gallo, D. A. (2012). Illusory expectations can affect retrieval-monitoring accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 391–404. doi:10.1037/a0025548

Metzger, R. L., Warren, A. R., Shelton, J. T., Price, J., Reed, A. W., & Williams, D. (2008). Do children "DRM" like adults? False memory production in children. *Developmental Psychology, 44*, 169–181. doi:10.1037/0012-1649.44.1.169

Moore, K. N., Lampinen, J. M., Gallo, D. A., Adams, E. J., & Bridges, A. J. (2017). Children's use of memory editing strategies to reject source misinformation. *Child Development, 89,* 219–234. doi:10.1111/cdev.12716

Moore, K. N., Lampinen, J. M., Gallo, D. A., & Bridges, A. J. (2017). Effects of feedback and test practice on recollection and retrieval monitoring: Comparing first graders, third graders, and adults. *Memory, 26,* 424–438. doi:10.1080/09658211.2017.1360356

Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1543–1547. doi:10.1037/a0020604

Raaijmakers, M. A., Smidts, D. P., Sergeant, J. A., Maassen, G. H., Posthumus, J. A., van Engeland, H., & Matthys, W. (2008). Executive functions in preschool children with aggressive behavior: Impairments in inhibitory control. *Journal of Abnormal Child Psychology, 36*, 1097–1107. doi:10.1007/s10802-008-9235-7

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.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: Some foundational issues. *Learning and Individual Differences, 7*, 145–162. doi:10.1016/1041-6080(95)90028-4

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217–236. doi:10.1068/p5117

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43*, 67–88.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1–24. doi:10.1006/jmla.1998.2611

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.

Strack, F., & Bless, H. (1994). Memory for nonoccurrences: Metacognitive and presuppositional strategies. *Journal of Memory and Language, 33*, 203–217. doi:10.1006/jmla.1994.1010

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, *16*, 75–84.

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). *Wechsler Intelligence Scale for Children: WISC-IV.* Paris: Les Editions du Centre de Psychologie Appliquée.

Appendix A

Table A. 1. *Examples of Stimuli Used in the Two Experiments. The English Translations of the French Words Are in Parentheses.*

Studied Items	Auditory Lures	Visual Lures	Semantic Lures	Unrelated Lures
Fourchette (Fork)	Fléchette (Dart)	Rateau (Rake)	/	Téléphone (Phone)
Bateau (Boat)	Gâteau (Cake)	/	Avion (Airplane)	Sablier (Hourglass)
Lune (Moon)	/	Banane (Banana)	Soleil (Sun)	Ecran (Sreen)
Louche (Ladle)	Loupe (Magnifying glass)	/	Cuillère (Spoon)	Bureau (Desk)
Domino (Domino)	/	Prise (Electrical outlet)	Toupie (Top)	Arbre (Tree)
Ballon (Ball)	Balcon (Balcony)	Orange (Orange)	/	Mouchoir (Handkerchief)

Appendix B

Table B.1. *Sensitivity (d') and Response Bias (C) Values Comparing Hits and False Recognitions for Each Type of Lure by Age Group for the Two Item Types in Experiment 1.*

	4–5 years (n = 14)		6–7 years (n = 14)		8–9 years (n = 14)		All (n = 42)	
	Picture	Word	Picture	Word	Picture	Word	Picture	Word
Visual/Auditory								
Sensitivity (d')	0.83 (.75)	0.32 (.87)	0.81 (.79)	0.72 (.68)	1.01 (.99)	1.21 (.67)	0.88 (.85)	0.75 (.81)
Response bias (C)	0.14 (.80)	0.18 (.79)	0.45 (.52)	0.24 (.66)	0.49 (.41)	0.26 (.46)	0.42 (.62)	0.23 (.64)
Semantic								
Sensitivity (d')	0.99 (.55)	0.52 (.67)	0.93 (.54)	0.47 (.80)	1.26 (.91)	1.09 (.72)	1.06 (.69)	0.69 (.77)
Response bias (C)	0.23 (.81)	0.28 (.82)	0.51 (.58)	0.11 (.68)	0.63 (.43)	0.20 (.57)	0.52 (.65)	0.20 (.68)
Unrelated								
Sensitivity (d')	0.83 (.75)	0.54 (.69)	0.81 (.79)	0.64 (.80)	1.21 (.78)	1.01 (.99)	0.94 (.85)	0.74 (.80)
Response bias (C)	0.30 (.86)	0.14 (.80)	0.45 (.52)	0.21 (.74)	0.52 (.41)	0.26 (.59)	0.43 (.62)	0.21 (.72)

Note. Standard deviations are in parentheses.

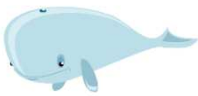
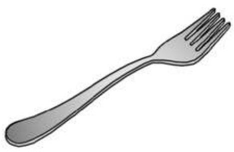
Table B.2. *Sensitivity (d') and Response Bias (C) Values Comparing Hits and False Recognitions for Each Type of Lure by Age Group for the Two Item Types in Experiment 2.*

	4–5 years (n = 14)		6–7 years (n = 14)		8–9 years (n = 14)		All (n = 42)	
	Picture	Word	Picture	Word	Picture	Word	Picture	Word
Visual/Auditory								
Sensitivity (d')	0.55 (.51)	0.50 (.54)	1.20 (.56)	0.83 (.56)	1.42 (.64)	1.20 (.33)	1.06 (.67)	0.84 (.56)
Response bias (C)	0.47 (.45)	0.48 (.46)	0.78 (.29)	0.87 (.29)	0.69 (.34)	0.73 (.33)	0.64 (.38)	0.69 (.40)
Semantic								
Sensitivity (d')	0.78 (.37)	0.50 (.54)	1.09 (.69)	0.83 (.56)	1.31 (.76)	1.20 (.33)	1.06 (.65)	0.84 (.56)
Response bias (C)	0.58 (.39)	0.48 (.46)	0.72 (.20)	0.84 (.29)	0.63 (.28)	0.73 (.33)	0.64 (.30)	0.69 (.40)
Unrelated								
Sensitivity (d')	0.63 (.48)	0.49 (.58)	1.23 (.59)	0.83 (.57)	1.48 (.69)	1.28 (.37)	1.11 (.68)	0.87 (.60)
Response bias (C)	0.50 (.42)	0.48 (.48)	0.78 (.27)	0.87 (.26)	0.71 (.31)	0.78 (.32)	0.67 (.35)	0.71 (.39)

Note. Standard deviations are in parentheses.

Appendix C

Table C.1. *Full List of Studied Words and Examples of Studied Pictures in the Two Experiments. The English Translations of the French Words Are in Parentheses.*

Studied pictures	Studied words
	Ampoule (Light bulb)
	Baleine (Whale)
	Ballon (Ball)
	Bateau (Boat)
	Bol (Bowl)
	Bouton (Button)
	Carotte (Carrot)
	Chaise (Chair)
	Chemise (Shirt)
	Ciseau (Scissors)
	Cloche (Bell)
	Collier (Necklace)
	Domino (Domino)
	Fourchette (Fork)
	Lion (Lion)
	Livre (Book)
	Louche (Ladle)
	Lune (Moon)
	Maison (House)
	Manteau (Coat)
	Oreille (Ear)
	Panier (Basket)
	Papillon (Butterfly)

Parapluie (Umbrella)

Pinceau (Paintbrush)

Poire (Pear)

Poubelle (Bin)

Salade (Lettuce)

Sifflet (Whistle)

Singe (Monkey)

Tabouret (Stool)

Tomate (Tomato)

Toupie (Spinning-top)

Trampoline (Trampoline)

Trompette (Trumpet)

Violon (Violin)