Developmental Science 7:2 (2004), pp 225–245

PAPER

Changes in reality monitoring and episodic memory in early childhood

Julia Sluzenski, Nora Newcombe and Wendy Ottinger

Department of Psychology, Temple University, USA

Abstract

The purposes of this research were to examine the developmental relation between reality monitoring and episodic memory, to link reality monitoring to autobiographical memory by using extended naturalistic events, and to examine prefrontal functioning as a potential contributor to development in reality monitoring and episodic memory. In Experiment 1, 4-year-olds were worse than 6- or 8-year-olds at reality monitoring after a week delay, despite the fact that they remembered more about real than imagined events and remembered different aspects of each. In Experiments 2 and 3, reality monitoring and episodic memory were evaluated for 4- and 6-year-olds immediately after the events occurred and, in Experiment 3, again after a week delay. Reality monitoring was at higher levels for both age groups, but age differences remained. These data suggest that preschoolers' difficulties with reality monitoring result from a combination of episodic memory deficits and strategic differences. In addition, correlation analyses more directly linked preschoolers' reality monitoring to episodic memory and supported the hypothesis that episodic memory development is related to prefrontal functioning.

Introduction

Accurate judgments concerning whether an action was performed or not (termed *reality monitoring*) constitute an essential part of healthy cognitive functioning (Johnson & Raye, 1981). Johnson, Hashtroudi and Lindsay (1993) proposed that reality monitoring judgments are part of a more general process of source monitoring, in which people evaluate the details of a memory in order to make decisions about its source (in this case, whether memory for an action has a real or an imagined source). In this paper, we refer to the details surrounding a memory as *memory characteristics*. There are different kinds of characteristics that may surround a memory, such as those that are perceptual (e.g. the colors of objects), contextual (e.g. the temporal order of events), semantic (e.g. apperceptive meaningful information about who and what was involved), cognitive/emotional (e.g. information about how one felt while one performed some action) or those that involve cognitive operations at the time of the memory formation (e.g. elaboration and organization of information). A premise of source monitoring theory is that real and imagined sources will be represented in memory by different amounts and kinds of characteristics. Real events are likely to be remembered with more detail overall than are imagined events. Furthermore, a memory for a real event is likely to contain a good deal of perceptual information compared to a memory for an imagined event, which will contain primarily information about cognitive operations involved in generating the imagination. The quantitatively and qualitatively different profiles of real and imagined sources serve as cues to discrimination between them.

Studies have shown that 3-year-olds have substantial difficulty with decisions about whether something actually occurred or not (Gopnik & Graf, 1988; Roberts & Blades, 1995; Taylor, Esbensen & Bennett, 1994; Welch-Ross, 1995).[1] Four-year-olds can succeed above chance at reality monitoring, although they perform less well than older children (Sussman, 2001). Six-year-old children perform at mature levels in judging whether *external* stimuli were actually encountered or merely imagined (e.g. whether they saw a movie or just heard about it) but

[1] Some of these studies used theory-of-mind tasks, which resemble source monitoring tasks in many ways and have been found to be statistically correlated with them (Leichtman, Morse, Dixon & Spiegel, 2000).

Address for correspondence: Julia Sluzenski, Temple Infant Laboratory, Haines House, 580 Meetinghouse Road, Ambler, PA 19002, USA; e-mail: jsluzens@temple.edu or Nora Newcombe, Department of Psychology, Temple University, 1701 N. 13th Street, Room 565, Philadelphia PA 19122-6085, USA; e-mail: newcombe@temple.edu

have more difficulty than older children and adults in judging *self-generated* actions (e.g. whether they combed their hair or just thought about doing so) (Foley, Durso, Wilder & Friedman, 1991; Foley, Johnson & Raye, 1983; Foley & Johnson, 1985). Age of success at reality monitoring most likely depends on a variety of task characteristics, such as the availability of kinesthetic cues to whether actions were performed (Foley, Aman & Gutch, 1987). Looking at children older than 6 years, researchers have found that improvement in reality monitoring continues at least to 10 years in certain situations (Parker, 1995; Sussman, 2001). In summary, existing data suggest that children begin to be able to make simple reality monitoring decisions at about the age of 4 years and that their ability to make such decisions increases from 4 years to at least 10 years.

Existing experiments have not, however, linked the advent or the development of reality monitoring abilities to the development of using memory characteristics to make reality judgments. Recall that Johnson *et al.* (1993) suggested that source monitoring is accomplished, in part, by an evaluation of the relevant memory characteristics. It seems clear that source monitoring would become increasingly difficult if such characteristics were impoverished or nonexistent. Poor memory, then, might be the cause of young children's difficulty with reality monitoring tasks. A point of clarification should be made regarding the existing literature on the relationship between memory for source and memory for the events. There is much evidence that memory for source is uncorrelated with memory for the events themselves (e.g. Foley & Johnson, 1985; Foley *et al.*, 1983; Lindsay, Johnson & Kwon, 1991), a finding that on the surface may seem to contradict the claims of source monitoring theory. However, 'memory for the events' has been defined in such studies as identification of which events took place, regardless of source, and which events did not take place. In this paper we were interested not in poor recognition memory for the events, but poor memory for the details that compose those events. To our knowledge no developmental studies of source monitoring have addressed such an issue. Also explored in this paper is an alternative pointed out by Parker (1995), namely, that young children have adequate memory to support reality monitoring but do not appreciate the fact that memory characteristics can be used to make source decisions. A large literature indicates that, while children as young as 2 years show certain strategies in memory situations (e.g. DeLoache, 1983), strategies for encoding and retrieval show considerable development across the early elementary school age range (Flavell, 1985). Thus, the first purpose of the present work was to gather data simultaneously on reality monitoring and memory characteristics in order to examine their relation developmentally.

We chose to pursue this purpose in the context of extended naturalistic events because we were also interested in linking reality monitoring development to autobiographical memory development. Adults can remember virtually nothing of their lives from the period prior to the age of 2; they remember more, but still comparatively little, from before the age of 5 (Bruce, Dolan & Phillips-Grant, 2000; Eacott & Crawley, 1998; Fivush & Hamond, 1990; Fivush, Hamond, Harsch, Singer & Wolf, 1991; Fivush, Hudson & Nelson, 1984; Usher & Neisser, 1993; Wetzler & Sweeney, 1986). This phenomenon, often called childhood amnesia, does not appear to be a matter of a deficit in memory in the most general sense, since implicit memory can be retained even when explicit memory is lacking (Newcombe & Fox, 1994). Nor does there appear to be a deficit in explicit memory in general, since a great deal of knowledge learned in early childhood is retained into adulthood (Wheeler, Stuss & Tulving, 1997). Indeed, there have been direct demonstrations of long-lasting explicit memory acquired during infancy using deferred imitation techniques (Bauer, Hertsgaard & Dow, 1994; Mandler & McDonough, 1995; Meltzoff, 1995). One conceptualization of the nature of childhood amnesia is as a deficit specific to episodic memory, including the self-related episodic memory that we call autobiographical memory (see Squire, 1992, for a discussion of the semantic–episodic memory distinction).

Previous studies of the development of reality monitoring have not studied events of the sort usually discussed by authors concerned with autobiographical memory. Typically, children have been brought into a laboratory and presented with words or sentences under various conditions or asked to perform or imagine simple actions (e.g. Foley, Harris & Hermann, 1994). Parker (1995) used naturalistic action vignettes, but the sequences were only four actions long (i.e. likely shorter than many everyday events) and mixed imagined and performed actions (in order to examine issues that occur with eyewitness testimony). Work with events consisting of a sustained series of interlinked actions, all imagined or all performed, would be more relevant to the issue of autobiographical memory. Such research has been conducted by Suengas and Johnson (1988) and Johnson, Foley, Suengas and Raye (1988) with young adults and by Hashtroudi, Johnson and Chrosniak (1990) with the elderly. Thus, a second purpose of the present work was to investigate the development of children's ability to decide if extended naturalistic events really happened or were imagined, in order to more directly link the findings from source monitoring studies to the development of autobiographical memory. The focus was on whether children could correctly judge whether a *whole sequence* of actions had been imagined or performed, rather than whether individual actions within a sequence had been imagined or performed.

Third, the studies aimed to examine preliminarily the hypothesis that the development of source monitoring and development of episodic memory are both linked to the development of prefrontal cortex. There are a number of reasons to suggest that this brain substrate may be involved in both source monitoring and episodic memory. Studies of patients with damage to prefrontal cortex have shown deficits in source memory (Janowsky, Shimamura & Squire, 1989; Schacter, Harbluk & McLachlan, 1984; Schacter, Kaszniak, Kihlstrom & Valdiserri, 1991) and in episodic memory (Wheeler, Stuss & Tulving, 1995). Similarly, elderly people who may suffer differential aging in prefrontal cortex have shown source deficits (Craik, Morris, Morris & Loewen, 1990; Glisky, Polster & Routhieaux, 1995; Parkin, Leng & Stanhope, 1988) and episodic memory deficits (Spencer & Raz, 1995). The role of prefrontal cortex in these abilities is also supported by the results of studies of normal adults using ERP and PET techniques. Prefrontal areas are activated during source tasks, and also during episodic and autobiographical memory tasks, perhaps especially when memory retrieval is effortful (e.g. Fink *et al.*, 1996; Nolde, Johnson & D'Esposito, 1998; Nyberg, Cabeza & Tulving, 1996; Ranganath & Paller, 1999, 2000; Raye, Johnson, Mitchell, Nolde & D'Esposito, 2000; Wilding & Rugg, 1996).

These findings are intriguing in light of evidence that frontal lobe development is an extended process in childhood, in studies of synaptogenesis (Huttenlocher, 1979, 1990; Huttenlocher & Dabholkar, 1997), in studies using PET (Chugani, Phelps & Mazziotta, 1987) and in studies using neuropsychological tests thought to be sensitive to prefrontal function (Becker, Isaac & Hynd, 1987; Chelune & Baer, 1986; Gerstadt, Hong & Diamond, 1994; Levin, Culhane, Hartmann, Evankovich, Mattson, Harward, Ringholz, Ewing-Cobbs & Fletcher, 1991; Passler, Isaac & Hynd; 1985; Welsh, Pennington & Groissier, 1991). Drummey and Newcombe's (2002) finding that 4-year-old children made extra-experimental errors in a source task far more frequently than 6- and 8-year-olds also speaks to this issue. In extra-experimental errors, the source of a fact is said to be someone other than the possible sources from the teaching session itself. Individuals suffering from prefrontal damage differentially show such errors (Schacter *et al.*, 1984; see related discussion of children's memory by Schacter, Kagan & Leichtman, 1995), and the fact that 4-year-olds also commit these errors suggests the possibility that relevant development in prefrontal cortex is incomplete at 4 years. In addition, there are correlations between executive function tasks thought to depend on prefrontal cortex (working memory and inhibition) and source memory in children aged 5 to 10 years (Ruffman, Rustin, Garnham & Parkin, 2001).[2]

In summary, the experiments reported in this paper were performed with three aims in mind. First, these experiments involved gathering data on memories of events at the time that reality judgments were made in order to study the developmental relation between memory and reality monitoring. Second, these experiments involved extended naturalistic events in order to explore the development of autobiographical memory. Third, children in these experiments were given a task believed to rely on prefrontal function in order to assess the hypothesis that reality monitoring and/or episodic memory are linked to prefrontal functioning in children, as has been found with adults. We chose category fluency, a task in which a person must generate examples of a specified category, as a measure of prefrontal function. While letter fluency is somewhat more frequently used to assess prefrontal function, it cannot be used with young children who lack phonemic awareness. There is evidence that category fluency is as affected by prefrontal damage as is letter fluency (Baldo & Shimamura, 1998), that it is correlated with prefrontal and temporal, but not hippocampal, gray matter volume in people with Alzheimer's (Fama, Sullivan, Shear, Cahn-Weiner, Marsh, Lim *et al.*, 2000) and that it is associated with prefrontal activity in PET studies of normal adults (Frith, Friston, Liddle & Frackowiak, 1991; Gourovitch *et al.*, 2000). Additionally,

[2] One should also consider the role that hippocampal areas play in source judgments and episodic memory. The role of the hippocampus in supporting formation of explicit memories is one of the best-documented facts in the neuroscience literature (Squire, 1992). One view of hippocampal function stresses its role in linking together distributed patterns of neuronal activation, that is, in binding together or associating information. Binding information to create richly textured memories is what may make memory episodic or, if self-related, autobiographical, and also allow for source judgments. There are various kinds of evidence for this hypothesis. For instance, certain kinds of episodic memory that require associations, notably verbal free recall and spatial memory, are more impaired by hippocampal damage than are other memory tasks (Henke, Kroll, Behniea, Amaral, Miller, Rafal & Gazzaniga, 1999). Similarly, children who suffer hippocampal damage in early hypoxic-ischaemic episodes have severe impairments of episodic but not semantic memory (Gadian, Aicardi, Watkins, Porter, Mishkin & Vargha-Khadem, 2000; Mishkin, Vargha-Khadem & Gadian, 1998). In addition, memory tasks requiring association or binding activate hippocampus in normal adults in PET studies when other memory tasks do not (Henke, Buck, Weber & Wieser, 1997). Ascribing a vital role to the hippocampus in episodic memory does not, however, exclude a role for prefrontal cortex. The hippocampus and the prefrontal cortex appear to work in concert on the task of supporting the binding together of various kinds of information (Mitchell, Johnson, Raye & D'Esposito, 2000; Mitchell, Johnson, Raye, Mather & D'Esposito, 2000).

category fluency was used as a measure of prefrontal function in a recent study of neuropsychological dysfunction in children with Down syndrome (Pennington, Moon, Edgin, Stedron & Nadel, 2003).

We performed two types of correlation analyses in order to explore the relationships among reality monitoring, memory and prefrontal function. The first involved correlating reality monitoring with memory measures and the second involved correlating category fluency with reality monitoring and memory measures. However, we reserve discussion of these correlations until after Experiment 3, when the small sample sizes from the separate studies can be pooled in meta-analyses aimed at detecting small or moderate effect sizes.

Experiment 1

This study explored in 4-, 6- and 8-year-old children the ability to distinguish between performed and imagined events, as well as the relation between this ability and the memory characteristics of the events. The experimental technique was adapted from the work of Hashtroudi *et al.* (1990), which examined the performance of younger and older adults who participated in four naturalistic events and imagined participating in four other such events. Events were activities such as packing a picnic basket or making a pot with clay, with each event containing a sequence of 18 to 20 individual actions. The following day subjects returned and were asked to write down as much as they could remember about each separate event. This task was used to assess the subject's memory characteristics for the events. Three weeks after the experiment subjects were given a reality monitoring task in which they were asked to determine whether each event had been imagined or perceived. On the recall test older adults reported fewer colors, objects, actions, references to nonvisual sensory information and spatial references than did the younger adults, but more thoughts, feelings and evaluative statements. Although both groups performed fairly well on the reality monitoring task, older adults had more difficulty than younger adults discriminating between imagined and perceived events.

Several basic characteristics of the paradigm were changed. First, cued recall was added as a measure of memory since free recall suffers from the drawback that participants of different ages might have different ideas about what should be included in their description of the event. To provide a complete measure of memory, recognition was also used to provide an upper-bound assessment of information retained. Second, the memory assessments and the reality monitoring questions were given in the same session rather than three weeks apart, in order to provide information about the status of the memory characteristics at the time the source judgments were actually being made. Third, children were read scripts to guide them through the event both when the event was enacted and when the event was imagined, rather than just the latter, in order to more closely equate the performed and imagined experiences. Last, some more minor aspects of the paradigm were changed to make the procedure possible with young children. Age-appropriate events were selected, four events were given instead of eight, and affective and cognitive operations at the time of the event were not queried due to the difficulty in verifying what children felt and thought.

Method

Participants

This study included a total of 60 children from three age groups, each containing 20 children: 4-year-olds (range = 3:6 to 4:4), 6-year-olds (range = 5:5 to 6:6) and 8-year-olds (range = 7:5 to 8:7). In this and in subsequent experiments, participants lived in the suburban districts of Philadelphia and were recruited from a commercially available mailing list containing birth date information.

Procedure

All participants were tested individually in a quiet room. However, most of the 4-year-old children were accompanied by their mothers in order to make them feel more comfortable. The study involved two testing sessions occurring exactly one week apart.

Session one

The first session included participation in the four events as well as administration of the vocabulary subtest from the age-appropriate Wechsler Intelligence Test. This task was included to rule out the possibility that differences in reality monitoring and/or episodic memory were related to differences in verbal IQ. Additionally, we wanted to rule out verbal ability as a potential confounding variable in the relationships among reality monitoring, episodic memory and category fluency. No mention of the subsequent memory test was made during the session.

Prior to starting the source task all children were given training, taken from a study conducted by Woolley and Bruell (1996), in understanding what it means to imagine and also in understanding the distinction between a real event and an imagined one. Children were first asked if they knew what it means to imagine something. Regardless of their responses to this question, they were told the following: 'Imagining is kind of like pretending. You just make a picture in your heads of what you are supposed to be imagining. Let's try it. Let's close our eyes and make a picture in our heads of an ice cream cone with two big scoops on it. Can you do that? Now, make a picture in your head that you are licking the ice cream real fast, before it drips.' After a few seconds had elapsed, the experimenter asked, 'So, what are you thinking about?' If the children failed to respond, the experimenter asked, 'Were you thinking about an ice cream cone?' Upon answering this question, the children were asked, 'What kind of ice cream cone were you thinking about?' This procedure served to get the children comfortable with closing their eyes and imagining events. Next, children were shown a book and instructed to open it and turn the pages one by one until they had finished looking at the book. Upon completion of this event, they were asked two questions: 'Did you actually lick an ice cream cone or just imagine licking the ice cream cone?' and 'Did you actually look at a book or just imagine looking at a book?' These questions were used to ensure that the children were capable of distinguishing between real and imagined events.

After the experimenter was certain that each child understood what it means to imagine something and was able to discriminate between real and imagined events, the study began. Participants were informed that a person on the tape recorder would be telling them exactly what to do in each event. They were also told to listen carefully to the person on the tape and to do only what the person on the tape was telling them to do. The children were encouraged to refrain from any talking during the events. The time period in between each of the events was used as a time when the experimenter and the child were able to talk. Participants were guided through each event, whether imagined or actually performed, by a detailed script presented on the tape recorder. The audiotape for all four events began with instructions to look at (or imagine) certain objects that were lying on the table in front of the child. The events were then narrated by explaining a sequence of actions.

The four event scripts were roughly equated in terms of the number of objects that were involved (5–8) and the number of individual actions within the sequences (18–20). The Appendix gives a full account of the event of planting a seed and the questions asked about it. The other three events were making pudding (pouring milk and pudding powder into a container, using a fork for mixing, adding peanuts, scooping the mixture into a mug, and sprinkling the top with chocolate chips), unpacking a picnic basket (taking out a plate and napkin, then taking out cheese, crackers and a banana and placing them on a plate, then taking out a cup, juice box and knife and placing them around the plate) and decorating a birthday invitation (taking a star-shaped paper, an envelope, a marker and a sticker out of a bin and placing them on a table, then coloring in a drawing on the star, placing a sticker at the bottom, tucking the paper into an envelope, sealing the envelope, and putting the marker back into the bin). Each event took approximately 5 minutes to enact. Events were always experienced in the following order: real, imagined, real and imagined. The four event scripts (making pudding, etc.) varied according to 12 possible orders, with each event appearing three times in each of the four positions; consequently, each event was imagined by half of the participants and performed by the other half.

Some children in all age groups occasionally needed the experimenter to stop the tape for a few seconds so that they would not fall behind in following the events. Although no data were collected on the frequency of such occurrences, it was the observation of the experimenters that, in general, more 6-year-olds than 8-year-olds and more 4-year-olds than 6-year-olds required pauses in the audiotapes. Also, younger children tended to need more nonverbal assistance (e.g. help putting on gloves) than did older children during such pauses. Despite these occurrences, though, all children were generally successful both in following the appropriate steps and in keeping pace with the audiotapes.

Following the four events, each child was given the vocabulary section of the age-appropriate Wechsler Intelligence Test. Children under 6 years of age were given the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) and children of age 6 or older were given the Wechsler Intelligence Scale for Children-III (WISC-III).

Session two

The second session took place exactly one week later. Session two included the source memory task, during which the children's memory for the events was tested, along with the administration of the category fluency task.

Participants were first reminded of their visit the previous week and were then told that they were going to be asked some questions about that visit. Participants were asked if they had performed or imagined a particular event (e.g. 'Did you actually unpack a picnic basket or just imagine unpacking a picnic basket?'). Second, children were asked to tell everything they could remember about the event (e.g. 'Tell me everything that you can remember about unpacking the picnic basket.'). Third, children were asked specific questions about the event (e.g. 'What color was the napkin?'). For any questions to which children did not respond or responded 'I don't know', a two-alternative forced-choice recognition question was asked immediately following the cued-recall question (e.g. 'Was the napkin red or blue?'). There were never any distractor recognition questions; the right answer always appeared as one of the two choices. All levels of the questions for one event were asked before proceeding to questions about another event. Participants' responses were recorded. Children were never corrected for wrong answers on any of these measures (nor were they in the subsequent experiments).

Following the memory questions, children participated in the category fluency task, in which they were asked to name as many members of a specified category as they could in one minute. This procedure took place for the category of 'animals' and then for 'fruits/vegetables'. Responses were recorded.

Scoring

Reality monitoring

A reality monitoring score was calculated by adding up the number of events correctly identified as having been either real or imagined. The highest possible score was a 4 and chance performance was a 2.

Free recall

One point was given for each piece of specific information recalled, with separate points given for mention of items, their shapes, colors and locations, for actions, and for indicators of temporal order (e.g. 'next' and 'before'). For example, if a child responded, 'There was a brown pot', he/she received two points for that answer. If the child said, 'First I picked up the brown pot', he/she received 5 points, one for indicating a temporal relation, one for the action of 'picking up', two for mentioning the brown pot, and one for mentioning that the pot was the thing picked up. Children did not receive points for any redundant information. For instance, if the child made such a statement, but had already mentioned that there was a brown pot, he/she would receive a point for mentioning that the pot was the thing picked up, but not two additional points for mentioning the brown pot again. In this manner, each of the four scripts was scored for a total possible score; maximum scores ranged from 60 to 70 points.

Cued recall

A total of 60 cued recall questions were administered, 15 questions for each event. There was only one possible correct answer for each question. Responses were scored as incorrect if either an incorrect answer was given, the child said 'I don't know' or no response was given. The final cued recall scores were calculated as the percentage of correct responses given to the questions.

Cued recall/recognition

The cued recall/recognition score comprised the total number of correct responses given either to cued recall questions or to the forced-choice recognition questions. The recognition questions were asked only if the child gave no response or said, 'I don't know' to a particular cued recall question. The recognition questions were not asked if the child gave an incorrect response to the cued recall question. The cued recall/recognition score was also calculated as the percentage of correct answers given.

Breakdown of question type

There were three types of questions asked: perceptual, spatial/temporal and semantic. It should be noted that it is difficult to draw definitive lines among the different question types. For instance, semantic questions can sometimes be answered using a perceptual image and perceptual questions with semantic knowledge about the event. In an attempt to limit this overlap, we defined each question type somewhat narrowly: perceptual questions were those referring to color or shape, semantic questions were those referring to objects or entities that could not be answered by remembering color or shape, and spatial/temporal as those referring to object locations or order of events. There was a total of 20 of each type of question asked, five of each type of question for each event. (See the Appendix for specific examples of each type from one of the events.) The percentage of correct responses was calculated for each type of question for both cued recall performance and the combined cued recall/recognition performance. Further, individual percentages for real and imagined events were also calculated.

Category fluency

The category fluency task was scored by counting up the number of correct items generated for both categories. Duplicate responses and basic level responses when subordinate responses were also given were not counted. For example, if bear, black bear, polar bear and koala bear were all given as responses, bear would not be counted. If, however, only bear had been given as a response, it would be counted.

Table 1 *Summary of means and standard deviations for 4-, 6- and 8-year-olds in Experiment 1*

	4-year-olds	6-year-olds	8-year-olds
Vocabulary	11.45 (2.33)	12.30 (2.58)	11.00 (2.96)
Category fluency	8.15 (3.42)	15.75 (5.70)	25.45 (7.32)
Reality monitoring	2.65 (1.04)	3.85 (.49)	3.80 (.52)
Free recall			
Real events	1.95 (1.90)	7.90 (4.22)	12.00 (5.24)
Imagined events	.55 (.76)	4.05 (3.59)	5.10 (3.51)
Cued recall			
Real events	19.00 (11.03)	52.83 (14.84)	64.00 (16.53)
Imagined events	5.17 (4.89)	20.83 (12.13)	32.33 (16.69)
Cued recall/recognition			
Real events	50.83 (12.79)	68.00 (10.62)	70.33 (14.79)
Imagined events	32.33 (15.60)	45.00 (12.59)	49.83 (18.75)

Note: The dependent variable in the vocabulary test is the standard score. The dependent variable in the category fluency task is the total number of items generated for both categories. The dependent variable in the reality monitoring task is the total number of events correctly identified as having been either real or imagined (out of 4). The dependent variable for free recall is the number of correct details of the events the child freely recalled. The dependent variable for both cued recall and cued recall/recognition is the percentage of correct answers (out of 30 for each event type in Experiments 1 and 2 and out of 12 for each event type in Experiment 3).

Results

In all analyses in this experiment and in Experiments 2 and 3, 'age' is treated as a between-subjects factor and all others as within. An alpha level of .05 was used for all statistical analyses, except for post-hoc tests, for which the appropriate Bonferroni adjustment was used on the alpha level. Means and standard deviations for all measures are presented in Table 1. A one-way analysis of variance (ANOVA) comparing the vocabulary scores for the 4-, 6- and 8-year-olds showed no significant difference across age groups, $F(2, 57) = 1.05$, $p = .36$. Thus, age differences in performance were not due to differences in verbal intelligence.

Reality monitoring

A one-way ANOVA showed that there were significant age differences in reality monitoring, $F(2, 57) = 17.34$, $p < .001$, with the 4-year-olds performing significantly worse than either the 6- or the 8-year-old children. The two older groups exhibited comparable levels of performance, close to ceiling. The 4-year-olds' performance, although low, was significantly better than chance, $t(19) = 2.80$, $p = .02$.

The 4-year-old children had substantial difficulty on the reality monitoring task. Of the 20 participants, only five were correct about all four events. Seven of the children said that they had imagined all of the events; one child said that he had actually performed all of the events. Of the remaining seven children, five said that they had imagined an event when in fact they had actually experienced it. The other two children said that they had participated in an event when they had only imagined it. Binomial tests confirmed that there were significantly more errors in which real events were reported as imagined rather than vice versa. In contrast to the performance of the 4-year-old children, only two of the 20 6-year-old children and three of the 20 8-year-old children made any errors at all.

Memory measures

There were three memory measures: free recall, cued recall and cued recall/recognition. Each of the measures should be interpreted keeping in mind the fact that each suffers from certain problems. For free recall, there are no specific cues given about what types of information to include; therefore, age differences could be attributed to different ideas about what details are important. There may also be age differences in willingness to talk to a strange adult and ability to verbalize information. Cued recall has the advantage of specifying required information and requiring only brief answers. However, children might remember certain aspects of the situation we chose not to probe, or age groups might differ in the willingness to venture an uncertain answer. For cued recall/recognition, there are two problems. First, the chance for a correct guess in the recognition task is 50%, and thus the measure is affected by variations in children's willingness to guess. Second, the recognition questions were only asked when children either did not respond to a cued recall question or said, 'I don't know'. Children who offered an incorrect answer that they were unsure about for a cued recall question were not given the chance to answer a recognition question.

Free recall

A 3 (age) × 2 (event type) ANOVA was performed, revealing a significant main effect of age, $F(2, 57) = 183.82$, $p < .001$, a significant main effect of event type, $F(1, 57) = 71.99$, $p < .001$, and a significant interaction between age and event type, $F(2, 57) = 11.12$, $p < .001$. All age groups had significantly more free recall for the real events than the imagined events. Comparisons between younger and older age groups were significant, except that 6- and 8-year-olds did not differ significantly for free recall of the imagined events. This lack of difference between 6- and 8-year-olds for free recall of the imagined events accounted for the age by event type interaction.

A possible question concerns whether children recalled actions or objects that they did not experience during the original events. In fact, intrusions were quite rare.

Table 2 *Mean percentages of correct responses for perceptual, spatial/temporal and semantic details of real and imagined events in Experiment 1*

	Real events	Imagined events
Cued recall		
Perceptual	50.50	20.33
Spatial/temporal	39.33	15.83
Semantic	45.17	22.33
Cued recall/recognition		
Perceptual	66.67	40.33
Spatial/temporal	59.00	40.00
Semantic	63.67	46.00

Although younger children produced incorrect information more often than did older children, the frequency with which any group did so was not high enough to warrant analyses. When such intrusions did occur, they almost always involved related rather than completely irrelevant information (e.g. reporting that a fork as well as a knife was removed from the picnic basket).

Cued recall

A 3 (age) × 2 (event type) × 3 (question type) ANOVA was performed on these data. There was a significant main effect of age, $F(2, 57) = 57.18$, $p < .001$, a significant main effect of event type, $F(1, 57) = 157.75$, $p < .001$, a significant main effect of question type, $F(2, 114) = 12.87$, $p < .001$, and an event type by question type interaction, $F(2, 114) = 3.68$, $p = .03$. There were significant differences in performance between all age groups for both types of events. All age groups had significantly better performance for real compared to imagined events. Responses were better for real events than for imagined events to all question types (perceptual, spatial/temporal, semantic). The interaction of event type and question type reflected the fact that more perceptual detail was remembered than spatial/temporal detail for real events, while more semantic detail was remembered than spatial/temporal detail for imagined events. This pattern shows that children of all ages, even the 4-year-olds, had some basis for making reality monitoring judgments (see Table 2, top panel).

Cued recall/recognition

A 3 (age) × 2 (event type) × 3 (question type) ANOVA revealed significant effects of age, $F(2, 57) = 12.64$, $p <$ .001, event type, $F(1, 57) = p < .001$, and question type, $F(2, 114) = 4.22$, $p = .02$. There were also interactions of age by question type, $F(4, 114) = 2.93$, $p = .03$, and event type by question type, $F(2, 114) = 3.58$, $p = .04$. Analysis of the interaction between age and question type showed that the 4-year-olds remembered about the same amount of the three kinds of information, while both 6- and 8-year-olds remembered significantly more semantic detail than spatial/temporal detail. Viewed another way, 4-year-olds remembered significantly fewer perceptual and semantic details than 6- and 8-year-olds, but for spatial/temporal questions there were no significant differences between age groups. The interaction between type of event and question type showed that, although all types of questions were more easily answered for real events as opposed to imagined events, for real events there was significantly more perceptual detail remembered than spatial/temporal and for imagined events there was significantly more semantic detail remembered than perceptual or spatial/temporal (see Table 2, bottom panel).

Discussion

The present study involved the comparison of 4-year-olds to older children in reality monitoring for extended sequences of naturalistic events after one week. The results showed that the performance of 4-year-olds was significantly above chance but also well below mature levels. What is the link between 4-year-olds' reality monitoring for events and their memory for specific characteristics of the events? Even the 4-year-olds seemed to have memories that potentially provided a basis for reality judgments, in that the children had different memory profiles for real and imagined events. Like the older children, 4-year-olds remembered more overall about real than imagined events by all three memory measures, including more perceptual, spatial/temporal and semantic information. They also remembered different kinds of information for each event type: for real events they remembered perceptual information best and for imagined events they remembered semantic information best. Remembering different amounts and kinds of memories from real and imagined events is consistent with reality monitoring studies with adults (Johnson *et al.*, 1988; Suengas & Johnson, 1988; Hashtroudi *et al.*, 1990). Thus, 4-year-olds had the potential basis for making accurate judgments.

If 4-year-olds have event memory profiles that are similar to those of older children and adults, why do they perform only slightly above chance at reality monitoring? One possibility that was discussed earlier is that 4-year-old children may suffer from a strategic difficulty. When deciding whether a source was real or imagined, they may not realize that real events are remembered better and with more perceptual detail than are imagined events. Another possibility is that memory may have been at such low levels for 4-year-olds after a week that there simply was not enough remembered detail to judge source with accuracy. For instance, impoverished perceptual and spatial/temporal information from the real events may have made them hard to distinguish from the imagined events, which likely had relatively little of such information from the start. The fact that errors were generally in the direction of saying that real events were imagined is consistent with this idea. Experiment 2 was designed to address these two possibilities for the 4-year-olds' difficulty with reality monitoring.

The finding that 4-year-olds misjudged real events more often than they did imagined events is inconsistent with evidence that children tend to make more errors in saying that a pretended or imagined event had actually been performed (e.g. Foley & Johnson, 1985; Parker, 1995; Sussman, 2001; Welch-Ross, 1995). We may have found a different pattern of errors merely by chance. However, a methodological difference between those studies and the present one raises a tentative but intriguing alternative. To date, reality monitoring studies with children have involved making judgments about the reality status of very simple events consisting of one action. In contrast, the present study involved reality monitoring of more complex events involving extended sequences. (Recall that each event took roughly five minutes to occur.) Reality monitoring of the latter kind of events may give rise to a different pattern of errors. In prior studies a common explanation has been that imagined or pretended actions may pose particular difficulty because of a paucity of cognitive operations or because the events gave rise to perceptual memories. Compared to single actions, however, sequences of many actions provide a richer set of cues with which to judge source. Consequently, for imagined events there may be much more information about cognitive operations than would be available for a single action, making it unlikely that extended sequences would be mistaken as real. If such a hypothesis is correct, then one would expect that, when children are tested on extended sequences after only a brief delay, errors would occur equally in both directions or possibly not at all. Children would be unlikely to decide that extended imaginations had really occurred, and real events would be remembered with enough detail to be easily distinguishable from imagined events.

Experiment 2

Experiment 2 examined whether low memory or poor strategy use was involved in the reality monitoring age differences found in Experiment 1. In the present experiment, memory measures were taken in the same session in which the events were performed or imagined rather than after a week delay. The purpose of immediate testing was to examine reality monitoring for events when memory performance would not be as low as it was in Experiment 1. If the poor reality monitoring of the 4-year-olds in Experiment 1 stemmed from the fact that memory overall was so impoverished, then remembering more event information would allow them to make more accurate reality monitoring judgments. However, if they have strategic difficulties in using memory characteristics to make reality judgments, they might still do poorly in reality monitoring even when memory is better.

Method

Participants

This study included two age groups consisting of 20 children each: 4-year-olds (range = 3:6 to 4:6) and 6-year-olds (range = 5:6 to 6:5).

Procedure

The only variation in procedure from Experiment 1 was that all measures were taken in one session. After the events, children were given about a 5-minute break during which they remained in the study room and were allowed to pick out stickers. Afterwards, all episodic memory measures were taken (reality monitoring, free recall, cued recall, and recall and recognition). The questions for each event were asked in the same order as the events took place. Last, both vocabulary and category fluency measures were taken. The session took approximately 40–45 minutes.

Results

Means for all measures are reported in Table 3. An independent samples *t*-test showed no significant age difference for vocabulary scores,[3] $t(37) = .035$, $p = .972$. Thus, any age-related differences were not due to group differences in verbal intelligence. With the extended length of the session, possible fatigue with the 4-year-olds was a concern. However, all children successfully made it to the end of the session. Furthermore, comparison of the vocabulary and category fluency scores of the 4-year-olds in Experiment 1 and Experiment 2 shows that they were roughly the same, indicating that performance was probably not affected by fatigue.

[3] Due to experimenter error, a vocabulary score was not obtained from one of the 4-year-olds.

Table 3 *Summary of means and standard deviations for 4- and 6-year-olds in Experiment 2*

	4-year-olds	6-year-olds
Vocabulary	12.57 (2.48)	12.55 (2.68)
Category fluency	9.32 (5.73)	14.85 (3.62)
Reality monitoring	3.65 (.81)	4.00 (0)
Free recall		
Real events	7.50 (4.64)	15.06 (5.79)
Imagined events	3.00 (3.57)	10.60 (6.79)
Cued recall		
Real events	49.33 (19.00)	81.50 (7.83)
Imagined events	27.83 (19.10)	57.17 (22.07)
Cued recall/recognition		
Real events	78.17 (8.83)	92.00 (7.30)
Imagined events	67.33 (16.77)	79.83 (12.90)

Note: For explanations of the dependent variables, see footnote to Table 1.

Reality monitoring

An independent samples *t*-test was used to examine age differences in performance on reality monitoring; results were marginally significant, $t(38) = 1.93$, $p = .062$. However, it should be noted that 6-year-old performance was literally at ceiling (none of the 6-year-olds made any errors), a situation that may have masked the size of the age effect. The 4-year-olds made errors, but, in contrast to Experiment 1, these errors were few in number and occurred equally often in both of the two possible directions. One child said that a real activity was imagined; one said that an imagined activity was real; one said that both real activities were imagined; one said that both imagined events were real and one real event was imagined.

Free recall

A 2 (age) × 2 (event type) ANOVA was used to examine the children's free recall performance. Results revealed a significant main effect of age, $F(1, 38) = 28.96$, $p < .001$, a significant main effect of type of event, $F(1, 38) = 31.43$, $p < .001$, and no significant interaction effect. Both age groups had significantly more free recall for the real events than for the imagined events, and 6-year-olds had higher free recall than did the 4-year-olds for both types of events.

Cued recall

A 2 (age) × 2 (event type) × 3 (question type) ANOVA was conducted to explore differences in cued recall performance. Results revealed a significant main effect of age, $F(1, 38) = 37.81$, $p < .001$, a significant main effect of event type, $F(1, 38) = 76.66$, $p < .001$, a significant main effect of question type, $F(2, 76) = 27.46$, $p < .001$, and a significant interaction between event type and question type, $F(2, 76) = 6.65$, $p = .002$. The 6-year-olds had higher cued recall than the 4-year-olds, both groups had higher cued recall for real events than for imagined events, and both groups had higher cued recall for semantic and perceptual questions than for spatial/temporal questions. Exploration of the interaction between event and question type revealed effects similar to those found in Experiment 1. For real events, cued recall score was higher for perceptual and semantic questions than for spatial/temporal questions. For imagined events, cued recall was higher for semantic than for perceptual, and higher for perceptual than for spatial/temporal questions.

Cued recall/recognition

A 2 (age) × 2 (event type) × 3 (question type) ANOVA was used to explore cued recall and recognition performance. Results revealed a significant main effect of age, $F(1, 38) = 16.05$, $p < .001$, a significant main effect of event type, $F(1, 38) = 37.81$, $p < .001$, a significant main effect of question type, $F(2, 76) = 16.29$, $p < .001$, and a significant interaction between event type and question type, $F(2, 76) = 16.06$, $p < .001$. The underlying differences accounting for both the main effects and the interaction effect mirrored those for cued recall, with the exception that, for imagined events, there was no difference between memory for perceptual and spatial/temporal questions.

Discussion

When memory was at higher levels than in Experiment 1, 4-year-olds' reality monitoring was much better, as can be seen by the fact that 75% of this age group made errors after a delay of a week and only 20% made errors when given a reality monitoring test immediately after the four events. Thus, a major reason for 4-year-olds' problems seems to be that when their memory has declined to relatively low levels, reality monitoring is difficult (as might be true for individuals of any age once memory is at low levels). The attenuated-memory hypothesis is also supported by the following finding: when 4- and 6-year-olds have similar levels of memory, they perform similarly on reality monitoring (see Figures 1 and 2 for comparison of the 4-year-olds in Experiment 2 to the 6-year-olds in Experiment 1 on reality monitoring and cued recall performance). Additionally, the fact that we did not find an error bias in Experiment 2 is evidence that extended sequences may not result in the same error

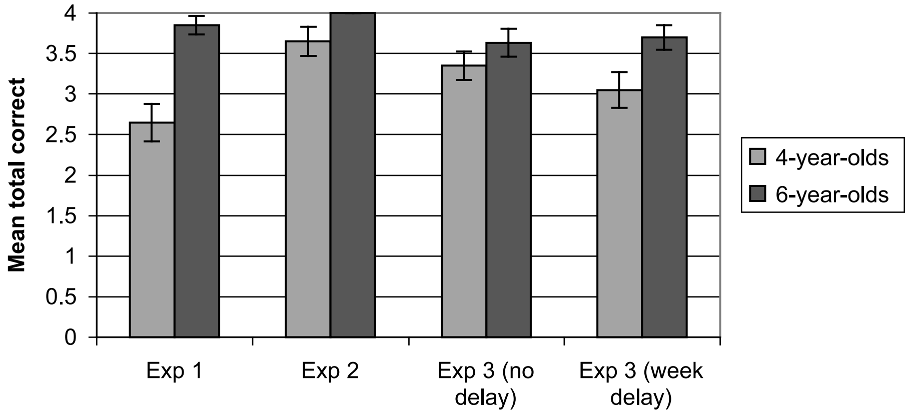

Figure 1 *Mean levels of reality monitoring for 4- and 6-year-old children across all experiments.*

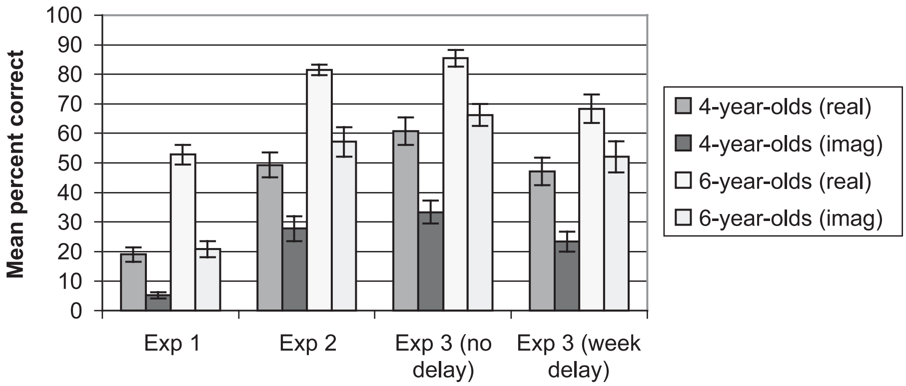

Figure 2 *Mean percent correct on cued recall of real and imagined events for 4- and 6-year-old children across all experiments.*

pattern as single actions do and that the direction of errors in Experiment 1 did, in fact, reflect impoverished memory levels. Memories may have been so bare after a week delay that most events seemed to have been imagined.

Despite this evidence that improved memory leads to better reality monitoring in preschoolers, 4-year-olds may still have a deficit in the strategy of using memory characteristics to judge reality status. Even though this age group had relatively high memory levels and even though they remembered very different profiles from the real and imagined events, they still performed slightly worse at reality monitoring than did older children within the conditions of Experiment 2. Furthermore, comparison of 4-year-olds in Experiment 2 to 6-year-olds in Experiment 1 does not rule out conclusively the possibility that there are strategic deficits. To make a stronger argument, it would be desirable to compare preschoolers and older children on reality monitoring when they have comparable memory levels *and* when they have endured the same delay between the events and the testing. We addressed this issue in Experiment 3.

Experiment 3

In this experiment, we again examined the effect of improved memory on children's reality monitoring, but now in the context of a week delay. In Experiment 2, memory questions were asked after a filled interval during which children experienced a combination of other events and/or questioning on other events. In addition, as in Experiment 1, children were not warned that there would be a memory test. In Experiment 3, we endeavored to maximize memory after a week delay, both by warning children that there would be a memory test and by asking the reality monitoring and memory questions immediately after each event. Such initial assessment would likely boost performance, a finding that would be consistent with behavioral and imaging data showing that memory retrieval tasks support renewed encoding of stimuli (Buckner, Wheeler & Sheridan, 2001) and with a study conducted by Parker (1995), which found that initial assessment improved children's memory after a delay. This design has two potential contributions. First, the experiment might serve to replicate the results of

Experiment 2 (i.e. high levels of memory in the preschoolers resulting in better reality monitoring), but with the more ecologically valid delay of one week. Thus, the design would make this work more relevant to the issue of autobiographical memory. In addition, the design might allow us to examine reality monitoring when the age groups have similar memory levels and when they have both experienced a delay of one week, by comparing 4-year-olds in Experiment 3 to 6-year-olds in Experiment 1. If 4-year-olds do in fact remember about the same amount as 6-year-olds in Experiment 1, but still perform worse at reality monitoring, then this would be further evidence that there are in fact age-related differences in the use of memory characteristics.

Method

Participants

This study included two age groups consisting of 20 children each: 4-year-olds (age range = 3:5 to 4:6) and 6-year-olds (age range = 5:4 to 6:7).

Procedure

Session one

The children participated in two sessions that were exactly one week apart. During the first session, children were brought into the study room and were given imagination training. The four events then took place. In an attempt to have the two groups perform as similarly as possible, two changes in procedure were introduced. First, before the events all children were told that they would be asked questions about what they remembered. Second, episodic memory measures were taken immediately after each event rather than after all of the events were completed. The measures were taken in the same order as in previous experiments: source identification, free recall, then cued recall and recognition. The cued recall questions were abbreviated so that instead of 15 questions for each event, there were only six (two semantic, two perceptual and two spatial/temporal). Children were not told that they would be asked questions on the following week.

Session two

Children were brought back one week later and taken into the study room. All episodic memory measures were taken again, with the same questions asked and in the same order in which they occurred in session one. Category fluency and vocabulary measures were also taken.

Table 4 *Summary of means and standard deviations for 4- and 6-year-olds in Experiment 3 (immediate test)*

	4-year-olds	6-year-olds
Vocabulary	11.60 (2.14)	11.80 (3.02)
Category fluency	7.65 (2.25)	15.50 (5.06)
Reality monitoring	3.35 (.81)	3.63 (.76)
Free recall		
Real events	6.85 (5.83)	14.26 (6.80)
Imagined events	4.75 (5.09)	12.68 (8.01)
Cued recall		
Real events	60.83 (20.75)	85.42 (12.67)
Imagined events	33.33 (17.08)	66.25 (16.58)
Cued recall/recognition		
Real events	78.33 (12.83)	93.33 (7.92)
Imagined events	60.42 (19.67)	81.25 (10.42)

Note: For explanations of the dependent variables, see footnote to Table 1.

Table 5 *Summary of means and standard deviations for 4- and 6-year-olds in Experiment 3 (repeated test)*

	4-year-olds	6-year-olds
Reality monitoring	3.05 (1.00)	3.70 (.66)
Free recall		
Real events	2.75 (3.97)	9.95 (5.96)
Imagined events	.90 (1.74)	6.70 (5.60)
Cued recall		
Real events	47.08 (21.00)	68.33 (21.75)
Imagined events	23.33 (15.17)	52.08 (24.00)
Cued recall/recognition		
Real events	72.08 (20.67)	80.83 (17.75)
Imagined events	54.17 (17.42)	70.42 (19.42)

Note: For explanations of the dependent variables, see footnote to Table 1.

Results

Means for all measures are reported in Tables 4 and 5.[4] Age groups did not differ in vocabulary scores, $t(38) = -.24$, $p = .8$, and so age-related differences could not be attributed to differences in verbal intelligence.

Reality monitoring

A 2 (age) × 2 (week) ANOVA was used to examine reality monitoring performance. Results revealed a marginally significant main effect of age, $F(1, 37) = 3.83$, $p = .058$, and no significant effects of week or interaction. For the 4-year-olds, errors in reality monitoring were quite common, with 12 of the children making at least one error in one of the sessions. The most common error was saying that a real activity had been imagined. When children were questioned immediately after the events,

[4] Due to experimenter error, reality monitoring and free recall data were not obtained from one of the 6-year-olds in session one.

there were nine errors of this type, and only two in the opposite direction (saying that an imagined event was real). When children were questioned again a week later, there were 12 errors of this type, and only three in the opposite direction. The difference in directionality was significant by binomial tests for both session one and session two.

Among the 6-year-olds, six children made at least one error in one of the sessions. Again, the most common error was in responding that a real activity was imagined. With immediate questioning, there were seven errors of this type, compared with none in the opposite direction. After a week, there were four errors of this type, compared with two in the opposite direction. The difference in directionality was significant by binomial test for session one but not session two.

Free recall

A 2 (age) × 2 (event type) × 2 (week) ANOVA revealed a significant main effect of age, $F(1, 37) = 27.31$, $p < .001$, a significant main effect of event type, $F(1, 37) = 7.83$, $p = .008$, a significant main effect of week, $F(1, 37) = 33.81$, $p < .001$, and no significant interaction effects. The 6-year-olds had higher free recall than did the 4-year-olds, both groups had higher free recall for real events than for imagined events, and both groups remembered significantly less during the second week than they did during the first week (exhibiting similar rates of forgetting).

Cued recall

A 2 (age) × 2 (event type) × 2 (week) × 3 (question type) ANOVA revealed a significant main effect of age, $F(1, 38) = 40.78$, $p < .001$, a significant main effect of event type, $F(1, 38) = 48.65$, $p < .001$, a significant main effect of week, $F(1, 38) = 38.55$, $p < .001$, and a significant 4-way interaction, $F(2, 76) = 4.78$, $p = .011$. In general, these main effects were due to the 6-year-olds having higher cued recall than did the 4-year-olds, both groups having higher cued recall for real events than for imagined events, and both groups having higher cued recall during the first week than during the second week. The 4-way interaction was not easily interpretable. As it did not approach significance in the analysis of cued recall and recognition (see below), most likely it is not a very robust finding.

Cued recall/recognition

A 2 (age) × 2 (event) × 2 (week) × 3 (question type) ANOVA was used to examine cued recall and recognition performance. Results revealed a significant main effect of age, $F(1, 38) = 18.17$, $p < .001$, a significant main effect of event type, $F(1, 38) = 31.54$, $p < .001$, a significant main effect of week $F(1, 38) = 39.73$, $p < .001$, and a significant interaction between event and question type, $F(2, 76) = 7.16$, $p = .001$. The 6-year-olds had higher cued recall and recognition than did the 4-year-olds, both groups had higher recall and recognition for real events than for imagined events, and both groups had higher cued recall and recognition during the first week than during the second week (again, exhibiting similar rates of forgetting). Exploration of the interaction between event and question type revealed that for real events more perceptual information was remembered than semantic, and for imagined events more semantic information was remembered than perceptual.

Discussion

Both age groups performed worse in reality monitoring in the initial test of Experiment 3 than they did in Experiment 2. These somewhat puzzling results were most likely due to the manipulation introduced in the third experiment of asking the reality monitoring questions immediately after each event. In effect, children did not experience both real and imagined events before they had to discriminate between them. Even though children in all experiments were given the imagination training before the start of the events, it is likely that experiencing all four events aided in their understanding of the difference between real and imagined. When children in Experiment 3 were asked to judge whether the first event had been real or imagined (the first was always real), they may have been confused by the question and guessed 'imagined'. Until one actually experiences the contrast between the event types, such a guess is probably a good one. After all, the experimenter and child did not really go on a picnic, or prepare a real birthday invitation, etc. The fact that the 6-year-olds also made such errors in the immediate test, but made none in the opposite direction, supports this interpretation of the findings. Furthermore, most of the children who made errors did so on the first real event but not on the second one (i.e. after they had also experienced an imagined event).

Memory levels a week after the events had been experienced were substantially higher in Experiment 3 than they were in Experiment 1. This enhanced memory, similar to an effect found by Parker (1995), was probably due to the fact that children in Experiment 3 had an opportunity to rehearse the events during the initial test. Consistent with this explanation is the fact that children who gave wrong answers to cued recall or recognition questions during the initial test frequently gave those same wrong answers during the repeated test, an effect also observed by Sussman (2001). This improved memory, as in Experiment 2, seemed to have a beneficial effect on reality monitoring performance. Even with the unexpected reality monitoring errors in session one, 4-year-olds performed better at the repeated test than they had in the delayed test of Experiment 1 (see Figure 1).

Nevertheless, despite the relatively good levels of memory seen in the second session of Experiment 3 and the improved reality monitoring, 4-year-old children still performed worse on reality monitoring than 6-year-olds did. Also, comparing across experiments using Figures 1 and 2, one can see that when 6-year-olds' and 4-year-olds' memory is approximately equal after a week delay, as for 6-year-olds in Experiment 1 and 4-year-olds in the repeated test of Experiment 3, it still appears that 6-year-olds do better than 4-year-olds at reality monitoring. It is unlikely that the different placement of the questions in the first session accounts for the age differences at the repeated test. By the second session the children had experienced both real and imagined events and so should have to some extent recovered from their confusion in the first session. In line with such an argument, 6-year-olds actually made fewer reality monitoring errors after a week delay than during the immediate test. Also, even with their confusions in the immediate test, 4-year-olds performed better in the repeated test than they did in Experiment 1, suggesting that they did to some extent benefit from our manipulations intended to bolster performance. The findings, then, support the notion that reality monitoring of preschoolers is impeded by strategic deficits as well as by memory deficits.

Correlations involving reality monitoring, memory measures and category fluency

Relationship between reality monitoring and episodic memory

The pattern of results across the three experiments indicates that children's reality monitoring was a function, in part, of their episodic memory for the events. However, we sought additional evidence of the relationship between reality monitoring and episodic memory by correlating the two variables within experiments and separately for the immediate and repeated tests of Experiment 3. For each age group we computed partial correlations between reality monitoring scores and total memory scores (i.e. the combined scores from real and imagined events) for each of the three memory measures (free recall, cued recall and cued recall/recognition), controlling for exact age and for scaled vocabulary scores.[5] The correlations were then combined across experiments in a small meta-analysis aimed at increasing statistical power to detect correlations. Combinations of the correlations involved transforming the Pearson's r values to Fisher's z values, finding the mean of the Fisher's z values, then transforming this mean into a Pearson's r, with degrees of freedom calculated from the sum of the sample sizes for the original r values. Two such analyses were done, one using the data from Experiment 1, Experiment 2 and the immediate test in Experiment 3, and the other using data from Experiment 1, Experiment 2 and the repeated test of Experiment 3. (Including both tests of Experiment 3 in the same analysis, when data were not independent, would inflate the results.)

Table 6 *Correlations between reality monitoring scores and memory measures, pooled across Experiments 1, 2 and 3, and controlling for age and scaled vocabulary scores*

	Using Exp. 3, immediate test		Using Exp. 3, repeated test	
	4-yr-olds	6-yr-olds	4-yr-olds	6-yr-olds
Free recall	−.08	.09	.06	−.02
Cued recall	.25*	−.13	.34**	−.13
Cued recall/recog.	−.09	−.04	.00	−.01

* $p < .05$; ** $p < .01$.

Results of these analyses are reported in Table 6. For the 4-year-olds, there was a significant correlation between reality monitoring and cued recall whether data from the immediate test or repeated test of Experiment 3 were included. This finding indicates that within experiments preschoolers' reality monitoring was related to the amount of information that they remembered about the events. Cued recall may have been the most appropriate measure in which to find a connection between reality monitoring and memory levels, since performance on cued recall was probably neither too effortful (such as in free recall) nor too automatic (such as in cued recall/recognition). In contrast to the results for the 4-year-olds, we did not find evidence of a relationship between reality monitoring and episodic memory for the 6-year-olds. The lack of evidence needs to be interpreted with caution, however, since reality monitoring in this age group was always at or close to ceiling. Thus, a relationship with episodic memory may not have been detectable in the present experimental setting.

[5] Since 6-year-olds performed at ceiling in reality monitoring in Experiment 2, it was not possible to compute any correlations involving reality monitoring for this experiment.

Relationship between category fluency and reality monitoring/episodic memory

With the data available from Experiments 1, 2 and 3, it was also possible to conduct a reasonably powerful test of the hypothesis that category fluency would be related to reality monitoring and/or to the three episodic memory measures. Partial correlations were computed separately for each age group in each experiment and separately for the immediate and the repeated tests of Experiment 3, again controlling for exact age and scaled vocabulary scores. The same meta-analytic procedures as described above were then used to combine the correlations.

If episodic memory is dependent on prefrontal areas to the extent that retrieval is effortful, as considerable evidence suggests, then one would expect significant correlations for category fluency and free recall (the most effortful memory task) and perhaps also for cued recall (the next most effortful), but not for cued recall/recognition. As shown in Table 7, correlations for category fluency and free recall were indeed significant for both 4- and 6-year-olds. Correlations of category fluency with cued recall were significant for 4-year-olds but not for 6-year-olds.[6] The nonsignificant finding for 6-year-olds can, however, be explained within the effortful-retrieval hypothesis, if one assumes that a cued recall task would involve less effort for older children.[7] The association between category fluency and effortful memory is probably not due to children with better vocabularies being more able to generate words and also being more able to engage in verbal recall, because it was evident with vocabulary held constant. Neither is the association likely to be due to more intelligent children having both better category fluency and better memory, because scaled vocabulary scores are a good proxy measure of general intelligence. The fact that the association survives partialling is consistent with the possibility that prefrontal areas are involved in both category fluency and in effortful retrieval.

Correlations between category fluency and reality monitoring were not significant. However, it would be premature to conclude that prefrontal function and reality monitoring are not related in development given the literature on this linkage in adults. It is possible, however, that reality monitoring is less directly related to prefrontal function than is episodic memory. In other words, reality monitoring may be related to prefrontal function to the extent that one recruits episodic memory when making a reality judgment. Additionally, reality monitoring at young ages may not be as determined by episodic memory processes as it is for more mature individuals because of children's difficulties in the strategic use of memory characteristics to make reality judgments.

Table 7 *Correlations between category fluency scores and memory measures /reality monitoring, pooled across Experiments 1, 2 and 3, controlling for age and scaled vocabulary scores*

	Using Exp. 3, immediate test		Using Exp. 3, repeated test	
	4-yr-olds	6-yr-olds	4-yr-olds	6-yr-olds
Free recall	.24*	.22*	.27*	.23*
Cued recall	.25*	.05	.34**	.00
Cued recall/recog.	–.10	–.17	–.02	–.27*
Reality monitoring	.10	–.14	.18	.04

* $p < .05$; ** $p < .01$.

General discussion

The three experiments in this paper demonstrate clearly that preschool children have problems distinguishing real events from events they have imagined. The stimuli used in the present studies were scenarios constructed to mimic event sequences that might actually occur in children's lives. One might have supposed that the ability to distinguish reality from imagination for extended events would be relatively easy, and hence evident at young ages. Such events would be engaging to children and would allow for a rich set of cues to their real versus imagined status. However, the data showed that 4-year-olds had substantial difficulty in judging whether complex extended events were experienced or not.

Could the age differences have resulted from the younger children not understanding what the task required? As discussed in relation to Experiment 3, the 'real' events were in a sense pretend events in that they took place in the context of play. In addition, the 'imagined' events in some sense actually happened – children were guided in their imagination by a taped story, rather than the children freely fantasizing in a way that they might during ordinary experiences of imagination. While this explanation is initially attractive, there are reasons to doubt its validity. First, the imagination training modeled after Woolley and Bruell (1996) was apparently effective. Few children, even among the youngest, seemed to experience any difficulty with imagining eating an ice cream cone or with answering immediate reality

[6] When data from the initial test of Experiment 3 were used, the correlations for 4- and 6-year-olds were not significantly different from each other, $z = 1.09$, $p > .05$. However, when data from the repeated test were used, the correlations did differ, $z = 1.89$, $p = .029$.

[7] There is, however, a surprising result in Table 6, namely the significant negative correlation of category fluency with cued recall plus recognition for 6-year-olds in one of the two analyses. This could be a chance result.

monitoring questions about whether they really ate it or only imagined eating it. Second, if young children simply had difficulty with the instructions, there would be no reason for their performance to be higher when their memory was tested more immediately rather than after a week – they would still be confused about the instructions. Third, the direction of the errors varied with delay. In Experiment 1, after a week delay, the 4-year-olds erred mainly by saying that real events had been imagined, but when tested after only a brief delay, in Experiment 2, they erred equally in both directions.

These experiments explored two hypotheses about why young children might have difficulty with reality monitoring. One possibility is that their memories are often relatively impoverished and that, in such a circumstance, it is hard to distinguish between real and imagined events using the richness of memories as a criterion for the decision. This hypothesis received support from the fact that 4-year-olds' reality monitoring benefited from more immediate testing and from repeated testing. Figure 1 summarizes the evidence on this point, by showing reality monitoring for 4- and 6-year-olds across the three experiments. Performance on reality monitoring clearly tracks the rise and fall of levels of memory (shown in Figure 2) across experiments. Additionally, the within-experiment correlations between reality monitoring and cued recall, when combined across experiments meta-analytically, revealed a reliable relationship for the 4-year-olds. This relationship was probably not found for the 6-year-olds due to their ceiling or near ceiling performance in reality monitoring throughout the experiments.

The second possibility is that 4-year-olds do not consistently appreciate that the richness of memories can be used to judge reality status, or that real memories have a predominance of perceptual detail while imagined memories have a predominance of semantic information and (presumably) information about cognitive operations. This hypothesis received support from the fact that age differences in reality monitoring were still evident in Experiments 2 and 3, when 4-year-old memory performance was improved from the low level seen in Experiment 1. Furthermore, even when age groups had comparable memory levels and had experienced the same delay (as was the case for the 6-year-olds in Experiment 1 and the 4-year-olds in the repeated test of Experiment 3), the 4-year-olds performed worse at reality monitoring. Further work could more specifically compare reality monitoring for different ages with manipulations designed to enhance memory for 4-year-olds and/or depress it for 6-year-olds.

These two possibilities are not mutually exclusive. It seems likely that both memory and judgment processes are relevant. Strategy appreciation and use may be present in a sufficiently robust form to allow for above-chance performance among 4-year-olds, and to allow for them to show biases such as thinking that real events are imagined when the memories of those events are quite faint. Yet, strategy use might be stronger and more consistent two years later. The factors may even grow in an inter-related fashion. As memory strengthens, strategies are easier to use. As strategies allow for distinctions between the real and the imagined, memories may seem clearer.

The fact that reality monitoring for complex extended events is difficult for preschool children supports a possible link between reality monitoring and autobiographical memory. What exactly is the nature of this linkage? The evidence presented here is that preschoolers, over time, have a tendency to judge that extended naturalistic events never occurred at all (Experiment 1). One consequence for autobiographical memory might be that young children never incorporate many personally experienced events into their autobiographical history. Such a hypothesis is consistent with the finding discussed earlier that the period of childhood amnesia extends to about 4 or 5 years of age, precisely the time when we observed a transition in the ability to judge whether a past event really occurred or not. Additionally, we found that questioning preschoolers after only a brief delay (Experiment 2) and questioning them repeatedly (Experiment 3) were two influences that seem to provide a buffer against losing real events as part of one's history. These findings are consistent with research showing that more elaborate parental questioning has been found to be associated with children's better recall of events (e.g. Reese, Haden & Fivush, 1993) and that children whose parents tend to engage them in discussions about the past can report earlier memories in adulthood (Mullen, 1994). Parental reminders and elaborations of the past most likely play a large role in the preservation of children's personally experienced events. Furthermore, the structure of preschoolers' narratives of events has been found to predict whether those events will be recalled later in life (Pillemer, Picariello & Pruett, 1994), suggesting that children's ability to discuss events (and, thus, rehearse them) is an important determinant in the preservation of autobiographical memories.

The finding in our studies that children were more likely to judge that a real event had been imagined rather than the reverse has important implications for children's eyewitness testimony. As previously mentioned, this finding is in contrast with other studies which have found that children are more likely to have difficulty in the reality monitoring of imagined events. For instance, both Parker (1995) and Sussman (2001) found that children showed a tendency to say that an imagined action had really occurred and did so whether tested after only a brief delay, after a delay of a week, or after a week delay when there had also been an initial test. Consequently, there may be a prevalent danger of children falsely accusing adults in certain situations, even when no misinformation has been suggested to the children (see Ceci & Bruck, 1993, for a related discussion on suggestibility effects). As hypothesized earlier, we may have found a different error pattern due to the extended nature of our events and the unlikelihood that an entire complex sequence that had been imagined could be mistaken as a real event. While our events lacked the interaction and emotional content of real-life abuse situations, they did resemble real-life events in the sense that they extended across time and involved a series of actions rather than single actions. Further studies should examine whether length of event (i.e. number of single actions within the event) is in fact an important determinant of the types of reality monitoring errors that children make. It may be true that, in the absence of immediate questioning, the failure to report a real event of abuse may be a more prevalent danger than false accusation.

The meta-analysis of correlations involving category fluency relates to the idea that prefrontal cortex may be involved in the encoding, binding and retrieval of different stimuli whose inter-relation allows for both episodic memory and reality monitoring. There is a great deal of evidence, reviewed in the introduction, supporting the involvement of prefrontal cortex in episodic memory and various kinds of source monitoring. In this study, the fact that category fluency was related to effortful retrieval, even with age and vocabulary controlled, offers support for the prefrontal hypothesis. Further work might establish the reliability of the finding that developments in episodic memory are related to changes in prefrontal function by using different measures for each of these variables. The relationship between reality monitoring and prefrontal function also needs to be explored further in order to reconcile the present null findings with findings from the adult literature. As argued earlier, this relationship may be more indirect than that between episodic memory and prefrontal function due to the fact that reality monitoring is to some extent epiphenomenal to episodic memory processes. Furthermore, it is possible that such a relationship is weaker in children than in adults because children may not evaluate and use episodic memory characteristics to the extent that adults do in making reality judgments.

It is possible that the significant correlations between episodic memory measures and category fluency were due to the involvement of medial temporal structures rather than of prefrontal cortex. For instance, category fluency has been shown to recruit medial temporal regions in addition to prefrontal regions in a functional magnetic resonance imaging study of normal adults (Pihlajamaki, Tanila, Hanninen, Kononen, Laakso, Partanen *et al.*, 2000). However, the fact that only the most effortful memory tasks were related to category fluency, a finding consistent with studies showing that effortful retrieval increases demands on prefrontal cortex (e.g. Nolde *et al.*, 1998), would be hard to explain with discussion of medial-temporal involvement. Also, extended development of prefrontal cortex relative to other brain regions, discussed in the introduction, renders a medial-temporal hypothesis less convincing than a prefrontal hypothesis. Additionally, Ruffman *et al.* (2001) found a relationship between source monitoring and prefrontal function using more widely accepted prefrontal tasks, offering support for the claim that changes in source monitoring are related to prefrontal development. Nonetheless, future work examining the relationships among prefrontal function, reality monitoring and episodic memory is needed to support the present prefrontal hypothesis.

Recently, there has been great interest in the role of prefrontal cortical development in the development of a variety of cognitive abilities. One domain that we have not mentioned so far is theory of mind, which has been suggested to have a prefrontal locus, perhaps because the tasks require inhibition of alternative prepotent responses (Carlson, Moses & Hix, 1998; Sabbagh & Taylor, 2000). Another domain is self knowledge. Various proposals exist concerning the relation of theory of mind and self knowledge to autobiographical memory (Nelson, 1993, 2000; Perner & Ruffman, 1995). One proposal is that thoughts and feelings about the self and others are intertwined with episodic memory in a bidirectional manner (as well as in terms of their neurological substrates, for a review of which, see Keenan, Wheeler, Gallup & Pascual-Leone, 2000). For instance, episodic memory and source monitoring are better for adults who are instructed to focus on speakers rather than on their personal reactions to speakers (Johnson, Nolde & De Leonardis, 1996). Consequently, children who are less able to understand others' perceptions and reactions may have lower levels of episodic memory than children who have this ability. Moving in the opposite causal direction, good levels of autobiographical recall might well support a better ability to reflect on the self and others. The intersections among cognitive, affective, social and neurological variables make this area a fascinating one for investigation.

Appendix

Sample script and questions for one event

I want you to imagine that you are going to plant a watermelon seed. Imagine that directly in front of you there is a rectangular placemat that has the alphabet on it. Also on the table is a brown pot to plant the seed in, a bag of dirt, a pair of gray gloves, a small, green, pointed gardening shovel, four small rocks, a white watering pitcher, and a watermelon seed. The first thing you are going to do is to imagine picking up the brown pot and putting it on the placemat with the alphabet on it. Next, imagine picking up the pair of gray gloves and putting them on. Then, pick up the small, green, pointed shovel. You are going to use the shovel to transfer dirt from the bag to the brown pot. So, put the shovel into the bag of dirt and scoop out some dirt and put it into the pot. Imagine doing this until you have filled the brown pot halfway full of dirt. Next, use the green, pointed shovel to dig a small hole in the middle of the dirt inside the pot. After digging the hole, imagine putting the shovel down inside the bag of dirt. Then, pick up the watermelon seed and put it in the hole that you just dug in the dirt. Use the other hand to push dirt into the hole around the seed until the hole is filled. Then, imagine picking up the white watering pitcher and pouring a small amount of water on the seed. Next, take off the gray gloves and place them across the top of the white watering pitcher. The last thing that you are going to do is to place the rocks around the seed to help protect the plant. So, imagine picking up one of the four rocks and placing it in the dirt around the seed. Imagine doing this with the other three rocks so that there is one rock on each side of the seed.

What did you use to dig the hole in the dirt? (semantic)
  Was it a shovel or your hands?
What kind of seed did you plant? (semantic)
  Was it a watermelon or a tomato?
What color was the pot? (perceptual)
  Was it black or brown?
What shape was the placemat? (perceptual)
  Was it round or rectangular?
Where did you put the pot? (spatial/temporal)
  Was it on the table or on the placemat?
What did you do first? (spatial/temporal)
  Was it put the pot on the placemat or put on the gloves?
Where did you put the shovel when you finished with it? (spatial/temporal)
  Was it on the placemat or in the bag of dirt?
What color were the gloves you wore? (perceptual)
  Were they gray or purple?
What shape was the shovel? (perceptual)
  Was it square or pointed?
Where did you put the gloves after you took them off? (spatial/temporal)
  Was it on the watering pitcher or in the bag of dirt?
What was the last thing that you did? (spatial/temporal)
  Was it put water in the dirt or put the rocks in the pot?
What color was the watering pitcher? (perceptual)
  Was it white or yellow?
What were the rocks used for? (semantic)
  Were they used for protection for the seed or to keep the dirt in place?
How many rocks did you put in the pot? (semantic)
  Was it three or four?
What was the design on the placemat? (semantic)
  Was it the alphabet or numbers?

Acknowledgements

Experiment 1 was part of a doctoral dissertation completed by the third author and Experiments 2 and 3 were the basis of a first-year project completed by the first author. We thank the Research Incentive Fund of Temple University for support, Camille Rocroi for help with data collection, and the children and their parents for participating. Portions of the data were presented at the Society for Research in Child Development, Albuquerque, New Mexico, April 1999.

References

Baldo, J.V., & Shimamura, A.P. (1998). Letter and category fluency in patients with frontal lobe lesions. *Neuropsychology*, **12**, 259–267.

Bauer, P.J., Hertsgaard, L.A., & Dow, G.A. (1994). After 8 months have passed: long-term recall of events by 1- to 2-year-old children. *Memory*, **2**, 353–382.

Becker, M.G., Isaac, W., & Hynd, G.W. (1987). Neuropsychological development of nonverbal behaviors attributed to 'frontal lobe' functioning. *Developmental Neuropsychology*, **3**, 275–298.

Bruce, D., Dolan, A., & Phillips-Grant, K. (2000). On the transition from childhood amnesia to the recall of personal memories. *Psychological Science*, **11**, 360–364.

Buckner, R.L., Wheeler, M.E., & Sheridan, M.A. (2001). Encoding processes during retrieval tasks. *Journal of Cognitive Neuroscience*, **13**, 406–415.

Carlson, S.M., Moses, L.J., & Hix, H.R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development*, **69**, 672–691.

Ceci, S.J., & Bruck, M. (1993). The suggestibility of the child witness: a historical review and synthesis. *Psychological Bulletin*, **113**, 403–439.

Chelune, G.J., & Baer, R.A. (1986). Developmental norms for the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, **8**, 219–228.

Chugani, H.T., Phelps, M.E., & Mazziotta, J.C. (1987). Positron emission tomography study of human brain functional development. *Annals of Neurology*, **22**, 487–497.

Craik, F.I.M., Morris, L.W., Morris, R.G., & Loewen, E.R. (1990). Relations between source amnesia and frontal lobe functioning in older adults. *Psychology and Aging*, **5**, 148–151.

DeLoache, J. (1983). Oh where, oh where: memory-based searching by very young children. In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 57–80). Hillsdale, NJ: Lawrence Erlbaum.

Drummey, A.B., & Newcombe, N.S. (2002). Developmental changes in source memory. *Developmental Science*, **5**, 502–513.

Eacott, M.J., & Crawley, R.A. (1998). The offset of childhood amnesia: memory for events that occurred before age 3. *Journal of Experimental Psychology: General*, **127**, 22–33.

Fama, R., Sullivan, E.V., Shear, P.K., Cahn-Weiner, D.A., Marsh, L., Lim, K.O., Yesavage, J.A., Tinklenberg, J.R., & Pferrerbaum, A. (2000). Structural brain correlates of verbal and nonverbal fluency measures in Alzheimer's disease. *Neuropsychology*, **14**, 29–40.

Fink, G.R., Markowitsch, H.J., Reinkemeier, M., Bruckbauer, T., Kessler, J., & Heiss, W.D. (1996). Cerebral representation of one's own past: neural networks involved in autobiographical memory. *Journal of Neuroscience*, **16**, 4275–4282.

Fivush, R., & Hamond, N.R. (1990). Autobiographical memory across the preschool years: towards reconceptualizing childhood amnesia. In R. Fivush & J.A. Hudson (Eds.), *Knowing and remembering in young children* (pp. 223–248). New York: Cambridge University Press.

Fivush, R., Hamond, N.R., Harsch, N., Singer, N., & Wolf, A. (1991). Content and consistency in young children's autobiographical recall. *Discourse Processes*, **14**, 373–388.

Fivush, R., Hudson, J., & Nelson, K. (1984). Children's long-term memory for a novel event: an exploratory study. *Merrill-Palmer Quarterly*, **30**, 303–316.

Flavell, J.H. (1985). *Cognitive development*. Englewood Cliffs, NJ: Prentice Hall.

Foley, M.A., Aman, C., & Gutch, D. (1987). Discriminating between action memories: children's use of kinesthetic cues and visible consequences. *Journal of Experimental Child Psychology*, **44**, 335–347.

Foley, M.A., Durso, F.T., Wilder, A., & Friedman, R. (1991). Developmental comparisons of explicit versus implicit imagery and reality monitoring. *Journal of Experimental Child Psychology*, **51**, 1–13.

Foley, M.A., Harris, J.F., & Hermann, S. (1994). Developmental comparisons of the ability to discriminate between memories for symbolic play enactments. *Developmental Psychology*, **30**, 206–217.

Foley, M.A., & Johnson, M.K. (1985). Confusions between memories for performed and imagined actions: a developmental comparison. *Child Development*, **56**, 1145–1155.

Foley, M.A., Johnson, M.K., & Raye, C.L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development*, **54**, 51–60.

Frith, C.D., Friston, K.J., Liddle, P.F., & Frackowiak, R.S.J. (1991). A PET study of word finding. *Neuropsychologia*, **29**, 1137–1148.

Gadian, D.G., Aicardi, J., Watkins, K.E., Porter, D.A., Mishkin, M., & Vargha-Khadem, F. (2000). Developmental amnesia associated with early hypoxic-ischaemic injury. *Brain*, **123**, 499–507.

Gerstadt, C.L., Hong, Y.J., & Diamond, A. (1994). The relationship between cognition and action: performance of children $3^1/_2$–7 years old on a Stroop-like day-night test. *Cognition*, **53**, 129–153.

Glisky, E.L., Polster, M.R., & Routhieaux, B.C. (1995). Double dissociation between item and source memory. *Neuropsychology*, **9**, 229–235.

Gopnik, A., & Graf, P. (1988). Knowing how you know: young children's ability to identify and remember the source of their beliefs. *Child Development*, **59**, 1366–1371.

Gourovitch, M.L., Kirkby, B.S., Goldberg, T.E., Weinberger, D.R., Gold, J.M., Esposito, G., Van Horn, J.D., & Berman, K.F. (2000). A comparison of rCBF patterns during letter and semantic fluency. *Neuropsychology*, **14**, 353–360.

Hashtroudi, S., Johnson, M.K., & Chrosniak, L. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging*, **5**, 119–126.

Henke, K., Buck, A., Weber, B., & Wiser, H.G. (1997). Human hippocampus establishes associations in memory. *Hippocampus*, **7**, 249–256.

Henke, K., Kroll, N.E.A., Behniea, H., Amaral, D.G., Miller, M.B., Rafal, R., & Gazzaniga, M.S. (1999). Memory lost and regained following bilateral hippocampal damage. *Journal of Cognitive Neuroscience*, **11**, 682–697.

Huttenlocher, P.R. (1979). Synaptic density in human frontal cortex – developmental changes and effects of aging. *Brain Research*, **163**, 195–205.

Huttenlocher, P.R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia*, **28**, 517–527.

Huttenlocher, P., & Dabholkar, A.S. (1997). Regional differences in synaptogenesis in human cerebral cortex. *Journal of Comparative Neurology*, **387**, 167–178.

Janowsky, J.S., Shimamura, A.P., & Squire, L.R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia*, **27**, 1043–1056.

Johnson, M.K., Foley, M.A., Suengas, A.G., & Raye, C.L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, **117**, 371–376.

Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3–28.

Johnson, M.K., Nolde, S.F., & De Leonardis, D.M. (1996). Emotional focus and source monitoring. *Journal of Memory and Language*, **35**, 135–156.

Johnson, M.K., & Raye, C.L. (1981). Reality monitoring. *Psychological Review*, **88**, 67–85.

Keenan, J.P., Wheeler, M.A., Gallup, G.G. Jr., & Pascual-Leone, A. (2000). Self-recognition and the right prefrontal cortex. *Trends in Cognitive Sciences*, **4**, 338–344.

Leichtman, M.D., Morse, M.B., Dixon, A., & Spiegel, R. (2000). Source monitoring and suggestibility: an individual differences approach. In K.P. Roberts & M. Blades (Eds.), *Children's source monitoring* (pp. 257–287). Mahwah, NJ: Erlbaum.

Levin, H.S., Culhane, K.A., Hartmann, J., Evankovich, K., Mattson, A.J., Harward, H., Ringholz, G., Ewing-Cobbs, L., & Fletcher, J.M. (1991). Developmental changes in performance on tests of purported frontal lobe functioning. *Developmental Neuropsychology*, **7**, 377–395.

Lindsay, D.S., Johnson, M.K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology*, **52**, 297–318.

Mandler, J.M., & McDonough, L. (1995). Long-term recall of event sequences in infancy. *Journal of Experimental Child Psychology*, **59**, 457–474.

Meltzoff, A.N. (1995). What infant memory tells us about infantile amnesia: long-term recall and deferred imitation. *Journal of Experimental Child Psychology*, **59**, 497–515.

Mishkin, M., Vargha-Khadem, F., & Gadian, D.G. (1998). Amnesia and the organization of the hippocampal system. *Hippocampus*, **8**, 212–216.

Mitchell, K.J., Johnson, M.K., Raye, C.L., & D'Esposito, M. (2000). fMRI evidence of age-related hippocampal dysfunction in feature binding in working memory. *Cognitive Brain Research*, **10**, 197–206.

Mitchell, K.J., Johnson, M.K., Raye, C.L., Mather, M., & D'Esposito, M. (2000). Aging and reflective processes of working memory: binding and test load deficits. *Psychology and Aging*, **15**, 527–541.

Mullen, M.K. (1994). Earliest recollections of childhood: a demographic analysis. *Cognition*, **52**, 55–79.

Nelson, K. (1993). Developing self-knowledge from autobiographical memory. In T. Srull & R. Wyler Jr. (Eds.), *The mental representation of trait and autobiographical knowledge about the self* (pp. 111–121). Hillsdale, NJ: Erlbaum.

Nelson, K. (2000). Memory and belief in development. In D.L. Schacter & E. Scarry (Eds.), *Memory, brain and belief* (pp. 259–289). Cambridge, MA: Harvard University Press.

Newcombe, N., & Fox, N.A. (1994). Infantile amnesia: through a glass darkly. *Child Development*, **65**, 31–40.

Nolde, S.F., Johnson, M.K., & D'Esposito, M. (1998). Left prefrontal activation during episodic remembering: an event-related fMRI study. *Neuroreport*, **9**, 3509–3514.

Nyberg, L., Cabeza, R., & Tulving, E. (1996). PET studies of encoding and retrieval: the HERA model. *Psychonomic Bulletin and Review*, **3**, 135–148.

Parker, J. (1995). Age differences in source monitoring of performed and imagined actions on immediate and delayed tests. *Journal of Experimental Child Psychology*, **60**, 84–101.

Parkin, A.J., Leng, N.R.C., & Stanhope, N. (1988). Memory impairment following ruptured aneurysm of the anterior communicating artery. *Brain and Cognition*, **7**, 231–243.

Passler, M.A., Isaac, W., & Hynd, G.W. (1985). Neuropsychological development of behavior attributed to frontal lobe functioning in children. *Developmental Neuropsychology*, **1**, 349–370.

Pennington, B.F., Moon, J., Edgin, J., Stedron, J., & Nadel, L. (2003). The neuropsychology of Down syndrome: evidence for hippocampal dysfunction. *Child Development*, **74**, 75–93.

Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, **59**, 516–548.

Pihlajamaki, M., Tanila, H., Hanninen, T., Kononen, M., Laakso, M., Partanen, K., Soininen, H., & Aronen, H.J. (2000). Verbal fluency activates the left medial temporal lobe: a functional magnetic resonance imaging study. *Annals of Neurology*, **47**, 470–476.

Pillemer, D.B., Picariello, M.L., & Pruett, J.C. (1994). Very long term memories of a salient preschool event. *Journal of Applied Cognitive Psychology*, **8**, 95–106.

Ranganath, C., & Paller, K.A. (1999). Frontal brain activity during episodic and semantic retrieval: insights from event-related potentials. *Journal of Cognitive Neuroscience*, **11**, 598–609.

Ranganath, C., & Paller, K.A. (2000). Neural correlates of memory retrieval and evaluation. *Cognitive Brain Research*, **9**, 209–222.

Raye, C.L., Johnson, M.K., Mitchell, K.J., Nolde, S.F., & D'Esposito, M. (2000). FMRI investigations of left and right PFC contributions to episodic remembering. *Psychobiology*, **28**, 197–206.

Reese, E., Haden, C.A., & Fivush, R. (1993). Mother–child conversations about the past: relationships of style and memory over time. *Cognitive Development*, **8**, 403–430.

Roberts, K.P., & Blades, M. (1995). Children's discrimination of memories for actual and pretend actions in a hiding task. *British Journal of Developmental Psychology*, **13**, 321–333.

Ruffman, T., Rustin, C., Garnham, W., & Parkin, A.J. (2001). Source monitoring and false memories in children: relation to certainty and executive functioning. *Journal of Experimental Child Psychology*, **80**, 95–111.

Sabbagh, M.A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: an event-related potential study. *Psychological Science*, **11**, 46–50.

Schacter, D.L., Harbluk, J.L., & McLachlan, D.R. (1984). Retrieval without recollection: an experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior*, **23**, 593–611.

Schacter, D.L., Kagan, J., & Leichtman, M.D. (1995). True and false memories in children and adults: a cognitive neuroscience perspective. *Public Policy and the Law*, **1**, 411–428.

Schacter, D.L., Kaszniak, A.K., Kihlstrom, J.F., & Valdiserri, M. (1991). The relation between source memory and aging. *Psychology and Aging*, **6**, 559–568.

Spencer, W.D., & Raz, N. (1995). Differential effects of aging on memory for content and context: a meta-analysis. *Psychology and Aging*, **10**, 527–539.

Squire, L.R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review*, **99**, 195–231.

Suengas, A.G., & Johnson, M.K. (1988). Qualitative effects of rehearsal on memories for perceived and imagined complex events. *Journal of Experimental Psychology: General*, **117**, 377–389.

Sussman, A.L. (2001). Reality monitoring of performed and imagined interactive events: developmental and contextual effects. *Journal of Experimental Child Psychology*, **79**, 115–138.

Taylor, M., Esbensen, B.M., & Bennett, R.T. (1994). Children's understanding of knowledge acquisition: the tendency for children to report that they have always known what they have just learned. *Child Development*, **65**, 1581–1604.

Usher, J.A., & Neisser, U. (1993). Childhood amnesia and the beginnings of memory for four life events. *Journal of Experimental Psychology: General*, **122**, 155–165.

Welch-Ross, M.K. (1995). Developmental changes in preschoolers' ability to distinguish memories of performed, pretended, and imagined actions. *Cognitive Development*, **10**, 421–441.

Welsh, M.C., Pennington, B.F., & Groisser, D.B. (1991). A normative-developmental study of executive function: a window of prefrontal function in children. *Developmental Neuropsychology*, **7**, 131–149.

Wetzler, S.E., & Sweeney, J.A. (1986). Childhood amnesia: an empirical demonstration. In D.C. Rubin (Ed.), *Autobiographical memory* (pp. 191–201). Cambridge: Cambridge University Press.

Wheeler, M.A., Stuss, D.T., & Tulving, E. (1995). Frontal lobe damage produces episodic memory impairment. *Journal of the International Neuropsychological Society*, **1**, 525–536.

Wheeler, M.A., Stuss, D.T., & Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychological Bulletin*, **121**, 331–354.

Wilding, E.L., & Rugg, M.D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain*, **119**, 889–905.

Woolley, J.D., & Bruell, M.J. (1996). Young children's awareness of the origins of their mental representation. *Developmental Psychology*, **32**, 335–346.

Received: 18 November 2002
Accepted: 12 May 2003