Journal of Experimental Psychology:
Learning, Memory, and Cognition
2006, Vol. 32, No. 1, 89–100

0278-7393/06/$12.00 DOI: 10.1037/0278-7393.32.1.89

Binding, Relational Memory, and Recall of Naturalistic Events: A Developmental Perspective

Julia Sluzenski
Vanderbilt University

Nora S. Newcombe and Stacie L. Kovacs
Temple University

This research was an investigation of children's performance on a task that requires memory binding. In Experiments 1 and 2, 4-year-olds, 6-year-olds, and adults viewed complex pictures and were tested on memory for isolated parts in the pictures and on the part combinations (combination condition). The results suggested improvement in memory for the combinations between the ages of 4 and 6 years but not in memory for the isolated parts. In Experiments 2 and 3, the authors also examined the developmental relationship between performance in the combination condition and free recall of a naturalistic event, finding preliminary evidence that performance on a memory task that requires binding is positively related to performance in episodic memory.

Keywords: episodic memory, binding, memory development, free recall

Binding refers to the process of encoding the relations among stimuli that co-occur (Cohen & Eichenbaum, 1993). Moscovitch (1994) used the term *cohesion* to describe the binding of stimuli and characterized the process as an early stage of consolidation that begins during the initial perception. This process is a critical component in the formation of complex memory representations. Take as a simple illustration the event of being introduced to a new person: One may form memories for the name and/or the face (the simpler stimuli); alternatively, and preferably for social reasons, one may form a memory for the relation between the name and the face (the more complex representation). This ability of binding enables the formation of memories of varying levels of complexity, from ones that are seemingly simple (e.g., memory for where in a list a certain word had appeared) to those that are quite rich (e.g., memory for all of the contextual details of one's wedding day).

The encoding of separate stimuli (e.g., color, object, or location) is a process that is separate from binding; these separate stimuli are believed to remain, at least at times, independently retrievable from the bound units. Memory for the simpler stimuli in the absence of memory for how those stimuli occurred together has been demonstrated in a variety of contexts. For example, people demonstrate retrieval of partial information about a word (e.g., the first letter of the word) during a "tip-of-the-tongue" experience (e.g., Brown & McNeill, 1966). Reinitz, Lammers, and Cochran (1992) examined the phenomenon of a "memory conjunction error," which refers to false memory for recombined parts of previously seen stimuli. When people studied compound words they sometimes recognized the proper base words but in the wrong combinations, showing that parts can be retrieved separately from the whole context (e.g., a person sees the words "jailbird" and "blackboard" and later reports having seen the word "blackbird"). Reinitz et al. (1992) also replicated these findings using faces instead of words as stimuli (see also Hannigan & Reinitz, 2000; Reinitz & Hannigan, 2001; Reinitz, Morrissey, & Demb, 1994). Recent research has shown that people sometimes incorporate memories from one event into their memories for a completely different event (Allen & Lindsay, 1998; Hannigan & Reinitz, 2003; Lindsay, Allen, Chan, & Dahl, 2004), demonstrating that these effects occur in more naturalistic situations and with events more complex than presentations of words and faces. Reinitz and Hannigan (2004) suggested that the mechanism of binding is necessary for separately encoded parts to be recalled later as a whole unit and, thus, for avoidance of partial and false memories (for related discussions, see Johnson, Hashtroudi, & Lindsay, 1993; Schacter, Norman, & Koutstaal, 1998).

The focus of this article is improvement in early childhood on a task that requires binding, as well as the developmental relationship between performance on such a task and episodic memory. Because it provides a potential basis for the construction of memories for highly complex events, many researchers believe binding to be a basic process in the formation of episodic memories (e.g., Cohen & Eichenbaum, 1993; Johnson & Chalfonte, 1994; Moscovitch, 1994; Shastri, 2002). In fact, Schacter et al. (1998) have argued that impairments in episodic memory tasks reflect disturbances in the ability to bind stimuli and/or in the subsequent reconstruction of bound information during recall. Johnson et al. (1993) have taken a similar approach in their explanation of source

Julia Sluzenski, Department of Psychology, Vanderbilt University; Nora S. Newcombe and Stacie L. Kovacs, Department of Psychology, Temple University.

This work was conducted as dissertation research by Julia Sluzenski. We thank the families of Brigantine (NJ) Elementary School, Egg Harbor Township (NJ) Education Station Preschool, and Temple Infant Laboratory (Ambler, PA) for their participation. We also thank Cristi Gallagher, Carly Slutzky and Meredith Meyer for their help with data collection and Thomas Shipley, Donald Overton, Karen Mitchell, Andrew Karpinski, Mark Wheeler, and Todd Jones for their helpful comments.

Correspondence concerning this article should be addressed to Julia Sluzenski, Department of Psychology, Vanderbilt University, Wilson Hall, 111 21st Avenue South, Nashville, TN 37203. E-mail: julia.sluzenski@vanderbilt.edu

monitoring errors, arguing that the binding of contextual details during the encoding of an event should facilitate accurate source memory for that event (see also Johnson & Chalfonte, 1994). Although the primary theoretical interest in this article is on binding (which begins during encoding), we acknowledge that memory performance also depends on retrieval or "reconstructive" processes noted by Schacter et al. (1998). Thus, throughout this article we use the term ""*binding* exclusively to refer to the mechanism in encoding that enables memory for the relations among stimuli and the phrase ""*relational memory* to refer to tasks that require binding (or to performance on such tasks).

Developmental investigations have shown that preschool children (between the ages of about 3 and 6 years) do not perform as well as older children in episodic memory tasks. For instance, Sluzenski, Newcombe, and Ottinger (2004) found that 4-year-olds and 6-year-olds differed dramatically in their memory for scripted events in which they had actively participated. Differences were consistent in tests of free recall, cued recall, and recognition, even when the events occurred just a few minutes before the memory tests. Although such results certainly do not rule out the possibility for age-related difficulties in retrieval, they do suggest that younger children's difficulties lie in part in poorer encoding of the events, given that age differences existed even with very brief delays between learning and test and even in relatively easy memory tasks (such as straightforward recognition tests). Consistent with an encoding hypothesis is evidence that working memory continues to improve well beyond the preschool years (e.g., Gathercole, Pickering, Ambridge, & Wearing, 2004), suggesting that younger children have more difficulty processing information for long-term storage. (Of course, working memory difficulties should also affect retrieval processes.)

Perhaps the most convincing evidence that childhood is a time of marked development in episodic memory draws from the source monitoring literature. In one of the earliest studies regarding this issue, Foley, Johnson, and Raye (1983) examined young children's ability to decide whether a memory originated from a private versus a public source (e.g., did I imagine myself making a statement or did I really make the statement?), a particular class of source discriminations called *reality monitoring*. These researchers found that 6-year-old children performed as well as adults when judging whether they or another person had made a statement and in judging which of two other persons had made a statement. However, children performed worse than adults in judging whether they had actually made a statement or whether they had only imagined making a statement (i.e., a reality status judgment). Research in the last two decades generally indicates that there is great improvement between the ages of about 3 to 6 years in the ability to source monitor in a variety of contexts (e.g., Drummey & Newcombe, 2002; Foley, Harris, & Hermann, 1994; Lindsay, Johnson, & Kwon, 1991; Roberts & Blades, 1995; Sluzenski et al., 2004; Sussman, 2001; Welch-Ross, 1995) and at least some continued improvement until 9 or 10 years of age (e.g., Drummey & Newcombe, 2002; Foley & Johnson, 1985; Lindsay et al., 1991; Markham, 1991; Parker, 1995).

Older adults also have difficulties in episodic memory tasks, such as those that involve source monitoring (e.g., Hashtroudi, Johnson, & Chrosniak, 1989, 1990; Hashtroudi, Johnson, Vnek, & Ferguson, 1994; Spencer & Raz, 1995). A few studies have shown that older adults perform worse than young adults in tasks that require binding, suggesting that their source monitoring difficulties may be due in part to age-related decline in binding. Chalfonte and Johnson (1996) asked young and older adults to learn a two-dimensional array of line drawings of common objects, each with a unique color and location. Participants performed a memory test for one of three features (i.e., the objects, colors, or locations), for the object–color combinations, or for the object–location combinations. Older adults performed worse than young adults in both combination conditions but not in the object or color conditions. Older adults also had worse memory for locations, but this age difference was not as great as that for the object–location combination condition. These findings support the hypothesis that older adults are impaired in relational memory even when memory for isolated features is relatively good (see also Mitchell, Johnson, Raye, & D'Esposito, 2000; Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000).

In the present research we examined whether children, like older adults, would have difficulty on a task that requires binding. One primary aim of Experiments 1 and 2 was to investigate the development of performance on such a task. A second aim was to test the hypothesis that development in the ability to remember the relations among stimuli contributes to episodic memory development. Specifically, we examined whether performance on a relational memory task would predict individual differences in free recall of a naturalistic complex event (Experiments 2 and 3).

We examined relational memory through a picture recognition task that closely resembled the task used by Chalfonte and Johnson (1996) in two important ways. These researchers compared memory for isolated features to memory for combined features when both abilities were tested in a similarly demanding paradigm—a recognition task. Participants also had to make combination judgments when the foils were new combinations of previously seen features. Therefore, participants could not respond correctly simply on the basis of discrimination between old and new features. We used a similar design in the present experiments: All conditions involved recognition judgments, and the combination condition involved new combinations of previously seen stimuli as foils.

However, modifications to the procedure used by Chalfonte and Johnson (1996) were needed to create a task that would be motivating and understandable enough to young children and still challenging enough for adults so that we could compare performance across groups. Experiment 1 was conducted as an exploration of a potential methodology in achieving this aim with 4-year-olds, 6-year-olds, and adults as the participants. In this experiment we attempted to maximize performance of the children: The experimenter gave participants explicit instructions on the nature of the task in the beginning of the session, and memory was tested after a delay of only a few minutes. In addition, we created stimuli that would be engaging for children: We used pictures of both common and uncommon animals, each appearing on a unique background. In a between-subjects design, participants made recognition judgments about the animals (animal condition), the backgrounds (background condition), or the animal–background com-

binations (combination condition).[1] Animals were paired with backgrounds arbitrarily to reduce guessing based on knowledge of the habitats of animals.

A related study has been conducted by Kirasic, Siegel, and Allen (1980), who asked preschool children, older children, and adults to study pictures of landmarks within their spatial contexts. In a forced-choice recognition task, participants had to select which landmarks they had seen and also select which landmarks in context they had seen. There were three conditions for the landmarks-in-context test: one in which foils were old landmarks in new contexts, one in which foils were new landmarks in old contexts, and one in which foils were new landmarks in new contexts. Younger children performed worse than older groups in all test conditions. This study suggests that preschool children have poorer memory than do older age groups for both the parts (suggested by the results of the isolated landmark condition) and for the relations among the parts (suggested by the results of the landmarks-in-context conditions). However, such a design is not ideal for answering the present question of development on tasks that require binding. First, preschoolers remembered fewer of the landmarks, which could have led to poorer memory for the landmarks in context. Second, there was no assessment of how well participants remembered the contexts alone. And finally, the foils in the landmarks-in-context conditions always involved a part of the picture that had never been seen. Thus, responding by all participants could have been based on discrimination between old and new rather than on memory for the original combinations of landmarks and contexts.

On the basis of prior literature showing that children improve on episodic tasks such as free recall and source monitoring between the ages of about 4 and 6 years, we expected there to be an age difference at least in the combination condition. In addition, we wanted to determine whether there is further refinement later in childhood, because the literature on source monitoring has shown that, at least in some situations, there is improvement after the preschool years (e.g., Lindsay et al., 1991). We also expected there to be a positive relationship between performance in the combination condition and performance on the free-recall task used in Experiments 2 and 3. We attempted to assess the size of this expected effect after controlling for the confounds of intelligence and precise age.

Experiment 1

Method

Participants

All age groups were drawn from suburban areas of Philadelphia; many of the adult participants were parents of the participating children. Parental consent was obtained for all of the children in this and in subsequent experiments. There were 140 participants: 48 four-year-olds (21 boys and 27 girls, mean age = 4.07 years, range = 3.42–4.50), 48 six-year-olds (17 boys and 31 girls, mean age = 5.97 years, range = 5.42–6.67), and 44 adults (14 men and 30 women, mean age = 30.86 years, range = 18–49). Participants were randomly assigned to one of three conditions. One third of each age group was in either an animal, a background, or a combination condition, with the exception of the adults, for whom there were only 14 in the animal and 14 in the background conditions.

In addition to these 140 individuals, there were 9 four-year-olds (6 in the combination condition and 3 in the animal or background condition) and 1 six-year-old (in the combination condition) who did not answer questions correctly during the training portion—even after demonstration by an experimenter that their answers were wrong—or who refused to participate in the session in its entirety. Data from these individuals were either not collected or not used because of incompletion.

Design and Procedures

Overview. For all participants, the experimenter first gave a demonstration of the task with explicit instructions for one of the three conditions. Participants were instructed to remember the animals, the backgrounds, or the animals together with their respective backgrounds. Following the training, the experimenter presented 16 pictures of animals in backgrounds; there was then a delay of about 2–3 min, followed by a test phase in which the participants made old–new responses to 16 test pictures. Below these procedures are described in further detail.

Training. Participants in all of the conditions examined the same four study pictures of animals in arbitrary backgrounds (e.g., a panda bear in the desert). The experimenter showed these pictures one at a time and asked the participants to remember either the animals, the backgrounds (also phrased as ""*places* for the children), or the animals together with the backgrounds. The experimenter pointed to the parts of the picture as she mentioned them. After showing each of the study pictures, she removed it from the sight of the participants.

Using four test pictures (showing them one at a time in a fixed order), the experimenter then asked participants to identify what they had just seen. To participants in the animal condition, the experimenter asked the following question: "Did you just see this animal? The rest of the picture is gone now, so it looks a little different." They responded to animals centered on a white page (order: old, new, old, new). To participants in the background condition, the experimenter asked, "Did you just see this background/place, but with an animal in it?" They responded to backgrounds without animals (order: old, new, old, new). To participants in the combination condition, the experimenter asked, "Is this where the animal was before?" Test pictures were both old and new combinations (order: new, old, old, new). Unlike later in the real test phase of the experiment, the new combinations during the training were not created by swapping the animals in two original pictures. Completely new backgrounds were used to reduce difficulty and maximize children's understanding of the question.

Children, particularly the 4-year-olds, sometimes gave incorrect responses during the training phase. If children in the animal or background condition gave an incorrect response, the experimenter showed the four study pictures again, then repeated the question. If children in the combination condition gave an incorrect response, the experimenter brought out the study picture with that animal, held it next to the test picture, and asked the question again. If children did not correct their answers, the experimenter ended the session.

Intentional learning phase. Immediately after the training, the experimenter told the participants that she would show them a lot of pictures of animals in different places and that the goal was the same as before (i.e., to remember the animals, the backgrounds, or the animals together with the backgrounds). Participants in all conditions saw the same set of 16 pictures, arranged in a fixed random order and displayed as pages in a book, one at

[1] The animal and background conditions correspond to the "feature" conditions of Chalfonte and Johnson (1996). However, the term *feature* is a problematic notion and perhaps even more problematic in terms of the present task, in which the "features" were rather complex stimuli. Consequently, the animals and backgrounds are hereafter referred to as the simpler "*units* or" *parts* of the pictures rather than as features, with acknowledgment of levels of binding that must also exist to form a memory of these units. The present work goes on the assumption, however, that the process of binding complex units is the same as that for binding simpler units.

a time, for 8 s each. Each picture contained a unique animal in a unique background (see Figure 1). During this time the experimenter commented on the pictures in a nonspecific way for the children (e.g., "These animals are in funny places, aren't they?"). These comments were made only to maintain children's orientation to the pictures and did not include labeling of things in the pictures. Although the youngest children needed more of such guidance, all age groups were generally successful at orienting to the pictures for the full 8 s.

Delay. There was a delay of 2–3 min before the test phase. During this time children picked out stickers or drew a picture; adult participants engaged in conversation with the experimenter.

Test phase. Participants in the animal condition made recognition judgments for 16 isolated animals centered on a white page (see Figure 2a), 8 of which had been shown during the presentation and 8 of which were new. Half of the participants saw one set of 8 original animals, and the other half saw the remaining original animals, resulting in two possible test sets. The same 8 new animals appeared as the new stimuli in each set. Participants in the background condition made recognition judgments for 16 isolated backgrounds (see Figure 2b); two test sets were created in the same manner as in the animal condition. Pictures in both conditions appeared in a semirandom order, with the restriction that no more than three old stimuli or three new stimuli appeared in succession.

The combination condition differed from the other conditions in that there were no new stimuli at test (see Figure 2c). There were two test sets: one using half of the original 16 pictures and the other using the remaining half. In each set, new combinations were created using the other half of the pictures. With these 8 pictures, four pairs were created, and then animals in each pair were swapped. Animals that were swapped were roughly equated in terms of size, and an animal assumed the same spot on the page that the original animal had occupied. Thus, approximately the same part of each background was occluded by the swapped animal. Pictures appeared in a semirandom order, with the restriction that no more than three correct combinations or three new combinations appeared in succession.

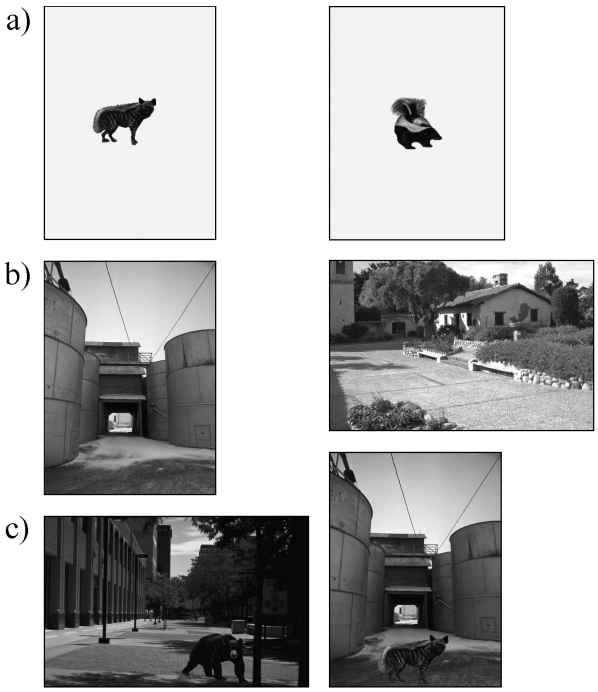

Figure 2. Examples of test stimuli for each condition. (Each picture was shown one at a time.) a: An old and a new stimulus in the animal condition. b: An old and a new stimulus in the background condition. c: An original combination and a new combination in the combination condition. (For any given animal, participants would see only the original or a new combination.)

Scoring

In this and in later experiments, we computed a sensitivity (d') score for each participant, separately for each condition.[2] In cases in which the proportion of hits was equal to 1, we adjusted the value to $1 - 1/(2N)$, where "" N is the maximum number of hits; in cases in which the proportion of false alarms was equal to 0, we adjusted the value to $1/(2N)$, where "" N is the maximum number of false alarms.

Results

There was a small subgroup of children who, even though they performed adequately during the training and made it to the end of the experiment, performed questionably during the test by giving either all "yes" or all "no" responses (3 four-year-olds from the animal condition, 4 four-year-olds from the background condition, and 1 six-year-old from the animal condition). Because such performance indicates poor understanding of a recognition task, data from these children were removed from the analyses.

Figure 1. Examples of animal–background combinations used during the learning phase.

Means for d' scores, proportions of hits, and proportions of false alarms are reported in Table 1. The primary focus for the analyses was on age-related performance in the combination condition compared with age-related performance in the other two conditions. In Experiments 1 and 2, we used planned contrasts to examine age differences in d' (at an alpha level of .05). Significance values are two-tailed, and effect sizes are reported as Cohen's d values. No corrections were made on the p values for this small set of planned contrasts. Making corrections in the animal or background condition would have reduced power to detect differences, and possible differences in these conditions were crucial for interpretation of performance in the combination condition.

Contrary to expectation, the 4- and 6-year-olds did not differ significantly in the combination condition ($p > .10$), although there was a moderate effect size (Cohen's $d = 0.56$). They also did not differ significantly in the animal condition ($p > .10$, Cohen's $d = 0.52$) or in the background condition ($p > .10$, Cohen's $d = 0.18$).

[2] For each participant, we also computed the criterion c (MacMillan & Creelman, 1991). With one exception of the adults in Experiment 2 exhibiting more liberal criteria than did children in both the initial and delayed background tests, there were no significant differences in c for any reported comparison.

Table 1
Experiment 1: Mean d′ Scores, Proportions of Hits, and Proportions of False Alarms

	4-year-olds		6-year-olds		Adults	
Condition and measure	*M*	*SE*	*M*	*SE*	*M*	*SE*
Animal						
d′	2.48	0.11	2.68	0.10	2.60	0.09
Hits	.88	0.03	.90	0.02	.93	0.03
False alarms	.06	0.03	.03	0.01	.07	0.02
Background						
d′	1.93	0.17	2.03	0.14	2.55	0.14
Hits	.71	0.05	.74	0.05	.88	0.03
False alarms	.06	0.03	.07	0.02	.04	0.02
Combination						
d′	1.34	0.23	1.78	0.17	2.28	0.20
Hits	.74	0.07	.88	0.03	.88	0.04
False alarms	.30	0.06	.27	0.04	.13	0.04

The difference between 6-year-olds and adults in the combination condition approached significance ($p = .09$, Cohen's $d = 0.62$). The two groups performed similarly in the animal condition ($p > .10$, Cohen's $d = 0.21$), but in the background condition, 6-year-olds performed worse than adults, $t(39) = 2.60$, $p = .01$, Cohen's $d = 0.96$.

We conducted an additional set of analyses exploring possible condition differences. For each age group, we examined differences among conditions with independent-samples *t* tests using a Bonferroni correction. These analyses revealed that both 4- and 6-year-olds did better in the animal condition than in the other two conditions and that adults performed comparably in all conditions.

Discussion

This methodology was generally successful for use with children, in that most children were able to comprehend the task and were motivated to complete it. The results provide only mixed support for the hypothesis that there is improvement from 4 to 6 years in performance on a relational memory task. Although there was a notable effect in the combination condition ($d = 0.56$), the sample sizes were not large enough for the effect to reach significance. In addition, there was a moderate effect in the animal condition ($d = 0.52$), raising the question of whether the difference in the combination condition was due to younger children's poorer memory for the parts of the pictures. However, if this argument were correct, then it would be difficult to explain why there was not a similar effect in the more difficult background condition. The greatest difference between the 6-year-olds and adults occurred in the background condition, which may have been due to children's interest in the animals (leading to a detriment in encoding of the backgrounds). Consequently, the difference that approached significance in the combination condition may have been due to poor memory for the backgrounds.

In making conclusions about performance in the combination condition, it would be useful to know whether 4-year-olds in the combination condition did not remember the animals well enough to support performance and also whether 6-year-olds in the combination condition did not remember the backgrounds well enough to support performance, questions that require a within-subject design to answer. Thus, in Experiment 2 we investigated the same age groups using the same general methodology but used a within-subject design, among other revisions.

Experiment 2

In Experiment 2 we incorporated a number of changes to the design in Experiment 1. Most important, we assessed memory for the parts and for the combinations within the same participants. In addition, we doubled the sample size per cell to increase power to detect differences in any of the three conditions. Finally, we made a set of changes to avoid the data exclusion that occurred in Experiment 1. There were two ways in which several 4-year-olds had been excluded from the sample: first, by never providing any or not providing enough data (these children were primarily in the combination condition; see the *Participants* section in Experiment 1) and, second, by giving all "yes" or "no" responses during the test (these children were all in the animal and background conditions; see the *Results* section in Experiment 1). In Experiment 2 the picture presentation involved the separate parts (the isolated animals and backgrounds) as well as the whole pictures. This change was made to avoid possible confusion of the later test question "Did you see this animal/background?" when, in fact, that animal or background had originally been part of a more complex picture. Also, we fine-tuned the training used in Experiment 1 to prepare children better for the later test. The changes in the learning phase and in the training had the added benefit of increasing the salience of the backgrounds relative to the animals.

Experiment 2 also included aspects that allowed us to examine the hypothesis that development in relational memory contributes to episodic memory development in early childhood. First, we included a measure of naturalistic episodic memory that involved listening to a story and then recalling it an hour later in a test of free recall. Second, the learning phase for the picture task was incidental, which allowed us to examine relational memory in a more naturalistic fashion. Also, there were two tests for the picture task conditions: a relatively immediate test (allowing for comparison to Experiment 1) and an hour-delayed test. The addition of a delayed test provided a measure of relational memory at about the same time of story recall (thus, better equating the memory demands of the two tasks). Finally, we administered the age-appropriate version of the Wechsler Intelligence Vocabulary

subtest (Wechsler, 1989, 1991) to control additionally for intelligence in this relationship.

Method

Participants

Children were from a school system in the suburban region of Atlantic City (NJ). Adults were students in introductory-level psychology classes at Temple University, participating for extra course credit. Participants were 32 four-year-olds (15 boys and 17 girls, mean age = 4.46 years, range = 4.01–4.97), 32 six-year-olds (15 boys and 17 girls, mean age = 6.41 years, range = 6.01–6.90), and 32 adults (7 men and 25 women, mean age = 23.43 years, range = 17–52).

Design and Procedures

Overview. Each participant experienced two sessions beginning 1 hr apart. Because the purpose of including the free-recall and vocabulary measures was to assess the developmental relationship between performance in the combination condition and free recall, we did not have the adults in Experiment 2 participate in these tasks. Thus, the sessions for the children were slightly longer than those for adults. However, the time increase for each session was only 5 min or less. (This time increase excludes that for collection of data on the vocabulary measure; however, this measure was taken in the second session after all other measures and therefore did not increase children's fatigue on these other measures.)

Before they viewed pictures in the first session, children listened to a brief story about a girl who bakes a cake on a rainy day. Adults, who did not participate in the free-recall task, first viewed pictures. For the picture presentation, the experimenter showed all participants a set of pictures that included isolated animals, isolated backgrounds, and animals in backgrounds, giving no instructions other than to look at the pictures. Immediately after this incidental learning phase was the training to prepare participants for subsequent testing. Finally, the experimenter tested them on their memory for the animals, the backgrounds, and the animal–background combinations, using only half of the test pictures during this first test. Approximately 1 hr after the start of Session 1, Session 2 took place. For the children, the experimenter first asked them to recall the story, tested them on the remaining half of the pictures, and then collected data on the vocabulary measure. For adults, this session consisted only of a test on the remaining half of the pictures. These procedures are explained below in detail.

Incidental learning phase for free-recall task (children only). The experimenter explained to the children that they would hear a story played on a tape. To encourage the children's attention, she told them to listen closely and think about whether they liked the story and whether they had ever heard it before. She did not tell them that memory for the story would later be tested. The experimenter then played the story (see the Appendix), which lasted approximately 2 min.

Incidental learning phase for picture task. The experimenter told participants that they were to look at pictures of animals. She then brought out a book that contained a set of 48 pictures, comprising 16 subsets. Each subset contained, in the following fixed order, an animal, a background, and the animal in that background. The sets themselves were ordered in a random sequence that was fixed for all participants.

The experimenter instructed the participants in the following manner: "I am going to show you a lot of pictures. What you will see is an animal, then a place where that animal likes to be—his favorite place, and then you will see that animal in his favorite place. You will see a whole bunch of pictures like that. All you have to do is look at them with me." The experimenter began displaying the pictures, turning the page every 5 s. In reference to the first picture, she said, "Here is the first animal." For the second picture, she said, "And look, here is his favorite place to be." For the third picture in the set, she said, "And here he is in his favorite place." In this manner, she proceeded to comment intermittently during the presentation.

Training on picture task. Immediately after the picture presentation, the experimenter began the training phase. She told all participants that she wanted to play a game with the pictures she had just shown them but first had to teach them the rules of the game. As in Experiment 1, the training involved study of combinations and then an immediate test. However, there were several notable changes made to the training used in Experiment 1. First, because of the within-subject design in Experiment 2, participants were trained on all conditions: They were instructed to remember the animals, the backgrounds, and the animal–background combinations. Second, after showing each study picture, the experimenter did not remove it from participants' sight but kept it on the table turned face down, forming a pile. This change was made to have a visual referent when asking the test questions (e.g., asking "Did you see this animal in here" while pointing to the facedown pile). Finally, the experimenter explicitly demonstrated a test picture that was a new combination of an old animal and an old background. Because of these alterations, it was necessary to make some changes both to the study and to the test pictures used for the training. Because participants were asked to study or remember more about the training pictures than in Experiment 1, only three animal–background combinations were used for the study and only two pictures per condition (six in total) were used for the test. The test pictures included, in the following fixed order: an old animal, a new animal, an old background, a new background, a new animal–background combination, and an old animal–background combination. Of importance is that the new animal–background combination animals involved both an old animal and an old background. When participants indicated that the animal was in the wrong place, the experimenter followed with "That's right. But what animal was there?" Regardless of the response, the experimenter brought out the study picture that showed that the background did in fact have a different animal in it. She noted explicitly that the animals had been switched and said, "You have to watch for when I try to trick you like that."

Initial test phase for picture task. After successful completion of the training, the experimenter told the participants that they would play the actual game. The 16 test pictures for each condition were identical to those used in Experiment 1, except that half of these pictures were used in the initial test and the remaining half in the hour-delayed test. The order of the two sets was counterbalanced across participants.

Participants were tested on the animal condition, the background condition, and the combination condition and always in that order. Placing the combination condition in either the first or second test position would allow additional exposure to the correct animals and backgrounds (i.e., exposure beyond the incidental learning phase). Because of the lack of counterbalancing, there was the possibility that any observed age difference in the combination condition would be due to younger ages fatiguing more quickly than older ages. However, this was an unlikely possibility, because in each session the testing for all three conditions took less than 5 min total.

The experimenter prefaced each of the three test portions with the instructions: "Some of the answers will be yes, and some of the answers will be no. So I want you to think really hard about what the answer is before you say it." These instructions were included to make the children sensitive to the presence of both old and new pictures in the test set and avoid the problem in Experiment 1 of perseveration in one response.

Free recall of story (children only). About 1 hr after the start of Session 1, the participants returned to the experimental setting. Adults were tested immediately on the remaining half of the pictures (see below). For the children, the experimenter reminded them of the story that they had heard on the tape. Once the children acknowledged having heard the story, the experimenter prompted them to tell everything about the story that they could remember. If the children did not respond, the experimenter encouraged them by saying that they could report anything at all and that it was okay if they could not remember everything.

Scoring of the free-recall test involved giving one point for each piece of specific information recalled, with points given for mentioning people, objects, locations, and actions. Points were also given for mentioning the relations among items. For example, if children responded, "First she baked a cake," they received six points for that answer: one for indicating a temporal relation, one for mentioning a girl, one for indicating her in relation to baking, one for mentioning the action of baking, one for mentioning a cake, and one for mentioning the cake in relation to baking. Note that this point system awards points not just for remembering separate items but also for remembering the relations among those items: There were three "item" points and three "relation" points in the example response above. Children did not receive points for any redundant information. For instance, if a child made such a statement but had already mentioned that there was a cake, the child would receive a point for mentioning that the cake was the thing baked but not an additional point for mentioning the cake again.

Because this scoring system awards points both for item responses and for relational information, we created a subset of scores that subtracted out the points for item responses. Our rationale for creating "relational-component" free-recall scores was to examine whether these scores would be more strongly correlated with performance in the combination condition.

Twenty-five percent of the free-recall responses were randomly selected for an intrajudge reliability check, and of the remainder of responses, 25% were randomly selected for an interjudge reliability check. In both cases, reliabilities were high (Pearson's *r*s = .97 and .99, respectively).

Delayed test phase for picture task. The experimenter told the participants that they were going to continue playing the picture game that they had played earlier. They were then tested on the remaining pictures in the same manner as during Session 1.

Vocabulary task (children only). The experimenter administered to children the age-appropriate version of the Wechsler Intelligence Vocabulary subscale. On this test, the children were asked to give the meanings of words in a list. The words were arranged according to increasing difficulty, and the test ended when the children reached the criterion number of failures. Children received a raw score that was then converted to a scaled score based on their exact ages and known distributions of scores.

Results

All children provided complete data in this experiment. In contrast to Experiment 1, all children demonstrated adequate performance during the training, and all cooperated during the remainder of the experiment. Also in contrast to Experiment 1, no children repeated all "old" or all "new" responses in the animal or background condition. A few children gave all "old" responses in the combination condition, but because these same children varied their responses in other conditions (thus, indicating that they understood the nature of the recognition task), they were not excluded from the analyses. Descriptive statistics for the picture task are in Table 2. As in Experiment 1, we used planned contrast analyses on the d' scores to make age comparisons.

Picture Task

In the initial test, 4-year-olds had lower d' scores than did 6-year-olds in the combination condition, $t(93) = -2.07$, $p = .04$,

Table 2
Experiment 2: Mean d′ Scores, Proportions of Hits, and Proportions of False Alarms

	4-year-olds		6-year-olds		Adults	
Condition and measure	*M*	*SE*	*M*	*SE*	*M*	*SE*
Initial test						
Animal						
d′	2.07	0.08	2.12	0.05	1.94	0.13
Hits	.93	0.02	.96	0.02	.91	0.03
False alarms	.04	0.03	.05	0.02	.08	0.03
Background						
d′	1.92	0.09	2.01	0.06	1.99	0.09
Hits	.86	0.03	.88	0.03	.94	0.02
False alarms	.04	0.02	.02	0.01	.09	0.03
Combination						
d′	1.01	0.15	1.40	0.12	1.53	0.12
Hits	.76	0.04	.79	0.04	.81	0.04
False alarms	.36	0.06	.20	0.05	.18	0.03
Delayed test						
Animal						
d′	1.86	0.10	1.86	0.09	1.97	0.11
Hits	.89	0.03	.86	0.04	.92	0.03
False alarms	.10	0.03	.06	0.03	.08	0.03
Background						
d′	1.62	0.12	2.01	0.07	1.83	0.08
Hits	.82	0.04	.88	0.03	.92	0.02
False alarms	.13	0.04	.02	0.02	.16	0.03
Combination						
d′	.84	0.12	1.03	0.14	1.12	0.14
Hits	.66	0.04	.68	0.04	.74	0.05
False alarms	.30	0.05	.27	0.04	.28	0.04

Cohen's $d = 0.52$, but the two age groups performed similarly in both the animal ($p > .10$, Cohen's $d = 0.10$) and background ($p > .10$, Cohen's $d = 0.20$) conditions. In the delayed test, the age difference in the combination condition diminished ($p > .10$, Cohen's $d = 0.26$), and there was still no age difference in the animal condition ($p > .10$, Cohen's $d = 0$). However, 4-year-olds remembered fewer backgrounds than did 6-year-olds in the delayed test, $t(47.84) = 2.80$, $p < .01$, Cohen's $d = 0.75$. (For this latter contrast, there was a violation of homogeneity of variance that required the use of a contrast that does not rely on this assumption.)

There were no differences between 6-year-olds and adults in the initial test in the combination condition ($p > .10$, Cohen's $d = 0.17$), in the animal condition ($p > .10$, Cohen's $d = 0.34$), or in the background condition ($p > .10$, Cohen's $d = 0.04$). In the second session, there were no differences in the combination condition ($p > .10$, Cohen's $d = 0.12$) or in the animal condition ($p > .10$, Cohen's $d = 0.20$). There was the surprising result that the 6-year-olds performed better than the adults in the background condition, although this difference did not quite reach significance ($p = .07$, Cohen's $d = 0.34$).

Exploration of condition differences within each age group, using Bonferroni-corrected dependent-samples t tests, revealed that in both the initial and delayed tests, all age groups performed significantly worse in the combination condition than in either the animal or background condition. This pattern of results differs from that in Experiment 1. Children's improved memory for backgrounds relative to combinations was probably due to changes in stimuli presentation to make the backgrounds more salient. Adults' poorer memory for combinations relative to animals and backgrounds may have been due to the change from incidental to intentional encoding.

Free-Recall Task and Correlational Analyses

As expected, 6-year-olds ($M = 18.34$, $SE = 1.42$) had significantly higher total free-recall scores than did 4-year-olds ($M = 5.91$, $SE = 1.28$), $t(62) = -6.52$, $p < .001$, Cohen's $d = 1.63$. Six-year-olds ($M = 8.66$, $SE = 0.82$) also had higher relational-component scores than did 4-year-olds ($M = 2.44$, $SE = 0.59$), $t(62) = -6.14$, $p < .001$, Cohen's $d = 1.53$. Twelve of the 32 four-year-olds did not recall any information, whereas all 32 of the 6-year-olds recalled at least some information. These findings mirror those of Sluzenski et al. (2004), who found marked differences between 4- and 6-year-olds in free recall of a single event that had occurred even in the very recent past.

For each age group, correlations were computed between d' scores in the picture task (separately for each condition and for initial and delayed tests) and free recall (separately for the total scores and for the relational-component scores). Each correlation was partialed for scaled vocabulary scores (to control for intelligence) and for exact age (to control for variables associated with age, such as size of vocabulary and willingness to talk to an adult). For 6-year-olds, the correlation between scores on the delayed test for backgrounds and the relational-component scores approached significance, $r(28) = .35$, $p = .06$. None of the other correlations approached significance for either age group (ps $> .10$).

The generally low level of 4-year-old performance, with 38% of this age group unable to recall any information, may have made this situation ill-suited for investigating relationships between free recall and the performance on the picture task. To address this issue, we conducted analyses again after excluding data from children who performed in the lowest 25th percentile for free recall (selected randomly among the 38%). These analyses yielded a positive relationship between performance in the delayed test for the combination condition and free-recall scores. Correlations approached significance, despite the sizable decrease in power after data exclusion: for total free-recall scores, $r(20) = .41$, $p = .06$; for relational-component scores, $r(20) = .40$, $p = .07$. Before data exclusion, these correlation values were .14 and .16, respectively.

Discussion

One primary finding from Experiment 2 was the improvement from age 4 to age 6 in the combination condition even though groups remembered the same amount of information in the animal and background conditions. Children did not know it was a memory test during the learning phase, did not know on what to focus, and were subsequently tested on all three conditions. Therefore, the age effect in the combination condition had to be due to a difference in relational memory and not to differential ability to study and/or remember larger amounts of information. It is interesting that, despite the many procedural differences between Experiment 1 and Experiment 2 (such as incidental vs. intentional encoding), the effect sizes for the combination condition comparison were approximately the same (Cohen's ds $= 0.56$ and 0.52, respectively).

The only difference between 4- and 6-year-olds that we found after an extended delay was in the background condition, in which 6-year-olds performed better than 4-year-olds. However, the fact that the 6-year-olds also showed a tendency to do better than adults in this condition points to an unusually elevated level for the 6-year-olds rather than to more rapid forgetting by the 4-year-olds. Age differences in the combination condition actually diminished somewhat from the initial to the delayed test, a finding that was most likely due to the difficulty of the task (see the General Discussion).

A second noteworthy finding was that 6-year-olds and adults performed comparably in almost all cases. The only exception was the nonsignificant trend of 6-year-olds performing better than adults in the delayed test for backgrounds. This finding is in stark contrast to the findings in Experiment 1, in which adults clearly performed better than the 6-year-olds in the background condition. The different findings may be due to a number of factors that varied from Experiment 1 to Experiment 2, such as the change from intentional to incidental learning, alterations in the presentation stimuli to make the backgrounds more salient, and differences in the participating adult populations (see the *Participants* sections of Experiments 1 and 2).

We found evidence in 4-year-olds that performance in the combination condition predicted episodic memory for a complex event, at least when delayed test scores for the combination condition were used in the analyses. There were several reasons for further investigation of the relationship. Because of the adult-level performance of the 6-year-olds in the combination condition, such a relationship may be quite small and hard to detect. Furthermore, the 4-year-olds' near-floor performance in free recall was undesirable, especially because poor performance may have been re-

Table 3
Experiment 3: Mean d′ Scores, Proportions of Hits, and Proportions of False Alarms

	Initial test		Delayed test	
Condition and measure	*M*	*SE*	*M*	*SE*
Animal				
d′	2.05	0.10	2.03	0.09
Hits	.93	0.04	.94	0.03
False alarms	.04	0.02	.07	0.03
Background				
d′	1.99	0.11	1.92	0.12
Hits	.88	0.05	.90	0.04
False alarms	.02	0.01	.08	0.02
Combination				
d′	.97	0.16	1.06	0.18
Hits	.83	0.04	.73	0.04
False alarms	.44	0.07	.29	0.07

lated not only to worse memory but also to greater shyness in younger children. This claim is supported by the fact that excluding data from children who recalled nothing in the free-recall test yielded a substantial positive relationship between performance in the delayed test for the combination condition and free recall.

Experiment 3 involved correlational analyses in 5-year-olds, with the rationale that they would have greater willingness and ability for verbal expression than would 4-year-olds. In addition, if there is improvement in relational memory from 4 to 6 years (as suggested by the results of Experiments 1 and 2) and if improvement in relational memory enables corresponding improvement in episodic memory, then one might expect there to be, in an intermediate age group, a sizable correlation between relational memory and episodic memory. A stronger relationship at 5 years would be consistent with findings in the literature that between the ages of 4 and 6 years there are drastic improvements in performance on memory measures, such as source monitoring and recall of complex events, with only more gradual improvements after the age of 6 (see the introduction).

Experiment 3

The goal of Experiment 3 was to examine in 5-year-olds the relationship between performance in the combination condition and free recall. It was expected that their performance in the free-recall task would be better than that of 4-year-olds, potentially creating a more favorable situation in which to assess the developmental relationship between the two measures.

Method

Participants

Children were from another school district of suburban Atlantic City (NJ). There were 24 five-year-olds (17 boys and 7 girls, mean age = 5.50 years, range = 5.02–5.94) who participated in this experiment.

Design and Procedures

All materials and procedures were identical to those in Experiment 2.

Results

Descriptive statistics for the picture task are shown in Table 3. Bonferroni-corrected dependent-samples *t* tests revealed that 5-year-olds performed significantly worse in the combination condition than in either the animal or background condition (in both the initial and delayed tests), similar to the performance of all age groups in Experiment 2. As expected, free-recall performance was in between that of the 4- and 6-year-olds ($M = 11.00$, $SE = 2.21$). However, 8 children (33%) did not recall any information. Therefore, we conducted the correlation analyses using the entire sample and then again excluding data from children in the lowest 25th percentile for free recall (selected randomly from the 33%). As in Experiment 2, correlations were computed between *d′* scores in each of the picture task conditions (separately for the initial and delayed tests) and free-recall scores (separately for total and relational-component scores). These correlations were then partialed for scaled vocabulary scores and exact age.

As with the 4-year-olds in Experiment 2, a positive relationship emerged between performance in the delayed test for the combination condition and free recall. This relationship was slightly stronger for relational-component scores than for total free-recall scores and also when the data from children in the lowest 25th percentile for free recall were removed (in all four cases, $ps < .05$). We report these partial correlations in Table 4, contrasted with the correlations between the other picture task conditions and free recall, all of which were nonsignificant and small in size ($rs < .20$).

Discussion

As expected, 5-year-olds showed intermediate performance in free recall of an event relative to slightly younger and older age groups, possibly creating a more favorable situation in which to examine correlations with this measure. Indeed, the correlation between performance on the combination condition (delayed test) and free recall was significant even after partialing out the variance associated with a measure of intelligence and with age. These results lend support to the claim that improved binding processes, and/or improved recall of the memory relations that result from

Table 4
Partial Correlations Between Free-Recall (FR) and Picture Task Conditions in Experiment 3 (5-Year-Olds)

	Total sample ($df = 20$)		Excluding 25% ($df = 14$)	
Session and condition	FR	Relational FR	FR	Relational FR
Initial test				
Animal	.13	.13	−.01	.03
Background	.11	−.02	.11	−.02
Combination	.09	−.18	−.13	.02
Delayed test				
Animal	−.05	.03	.04	.13
Background	.13	.16	.09	.13
Combination	.42*	.46*	.55*	.61**

* $p < .05$. ** $p = .01$.

binding, contribute to preschool development in naturalistic episodic memory.

General Discussion

Preschool Development in Relational Memory

This research demonstrates improvement in early childhood on a task that requires binding. In Experiment 1, which involved an intentional learning situation and a between-subjects design, there was a tendency for preschool children to remember combinations of stimuli more poorly than older children (although, with the small sample sizes, this effect did not reach significance). There was also a nonsignificant but similar age difference in the animal condition, leaving unanswered the question of whether preschool children have worse memory for the relations or worse memory for all stimuli (implicating a more general encoding difficulty). The results of Experiment 2, which involved an incidental learning situation and a within-subject design, demonstrated that preschool children do perform worse than older children in a condition that requires binding. The same 4-year-olds who performed as well as 6-year-olds in the animal and background conditions performed worse than 6-year-olds when trying to recall which animals and backgrounds had occurred together. The similar effect sizes observed for the combination condition in Experiment 1 and for the combination condition (initial test) in Experiment 2 suggest that the age difference may be present regardless of whether the learning is incidental or intentional.

The age difference in the combination condition grew smaller over time in Experiment 2, although the arbitrary nature of the animal–background pairings may have made these relations exceptionally difficult to remember. In addition, comparison across Tables 1 and 2 suggests that the switch from intentional to incidental learning led to lower memory performance. Because learning involved arbitrary relations and was incidental, the memories of even the older age groups may have been too fragile to survive the interval of 1 hr.

Later Development

The results of Experiment 2 suggest that relational memory may be near or at adult levels by about the age of 6 years. This finding is consistent with research indicating that the number of one-time events that adults can remember from childhood increases sharply for events that occurred after 5 or 6 years of age (e.g., Pillemer & White, 1989). With inadequate ability to encode relations and/or to recall those relations, younger children would lack the basis essential for remembering single events and probably necessary for rich autobiographical memory. The results of Experiment 2 are slightly inconsistent with source monitoring studies, which have pointed to a more protracted development of episodic memory (as discussed in the introduction). However, source monitoring is a complex ability clearly dependent both on the binding and recall of relations and on complex reasoning strategies for which adults most likely have an advantage (Johnson et al., 1993).

Developmental Relationship Between Relational Memory and Episodic Memory

Experiments 2 and 3 provided evidence for the hypothesis that improvement in relational memory leads to changes in episodic memory in early childhood, as indicated by the positive correlations between performance in the delayed test of the combination condition and free-recall scores. In 5-year-olds, these correlations were significant and of moderate size even after controlling for exact age and intelligence. The correlations were stronger when data were used from children in the upper 75th percentile for free recall (eliminating data from children who, because of either poor memory or shyness, received a score of 0 on the free-recall test). The relationship was also stronger when the free-recall scores did not include item information, perhaps making the scores more sensitive to relational memory.

Correlations in the 6-year-olds may have been smaller than in the 5-year-olds because of a diminished relationship by this point in development, when children are reaching an adult level in relational memory. The initial null results in 4-year-olds were probably related to a floor effect in free recall. Supporting this explanation is the fact that a notably stronger relationship existed when a quarter of the sample was removed for receiving a score of 0 in the free-recall test. It is also very possible that the relationship was not easily detectable in the current experimental design. We used a single relational memory condition and a single episodic memory task; this limited manner of testing does not necessarily lead to a good estimate of either ability. A better approach would be to use a battery of tests measuring both abilities and to use composite scores when computing correlations (e.g., Glisky, Polster, & Routhieaux, 1995).

The possibility exists that there are other variables that could account for the relationship, such as attention and motivation. Yet another possibility is that performance in the combination condition was correlated with free recall because both measures depend on memory for simpler stimuli (or "item" memory). However, according to either of these accounts, one would expect that free recall would also correlate with performance in the animal and background conditions. In fact, these correlations were quite small (see Table 4).

Encoding Versus Retrieval Hypotheses

In the introduction we noted that binding is only one determinant of performance on relational memory tasks. An important avenue for investigation will be to determine the exact nature of the developmental changes that lead to better performance. We do not know from these studies whether 4-year-olds do not encode properly the relations among stimuli (a binding failure) or whether they initially encode but have subsequent difficulty in the retrieval of those relations.

In the adult literature, it has been hypothesized that memory conjunction errors occur because of a failure in binding processes, both in normal adults (Reinitz & Hannigan, 2004; Reinitz et al., 1992, 1994) and, with greater frequency, in amnesia patients (Chalfonte, Verfaellie, Johnson, & Reiss, 1996; Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996; Reinitz, Verfaellie, & Milberg, 1996). According to an encoding account of the present findings, younger children should perform worse than older children in the combination condition because of poorer binding between animals and backgrounds. Also according to this account, performance on tasks that require binding should correlate with episodic memory because better binding enables better episodic encoding.

Alternatively, younger children may have greater difficulty than do older children at the time of retrieval. Some researchers have hypothesized that a recall-to-reject mechanism enables resistance to a false alarm for a new item that shares characteristics with an original item (e.g., Hintzman & Curran, 1994; Rotello, MacMillan, & Van Tassel, 2000). Specifically, when a person encounters such a lure, he or she may recollect the original item that was similar to the lure and, consequently, avoid a false alarm. This process requires that one simultaneously consider both the lure and the original stimulus. Mitchell, Johnson, Raye, Mather, & D'Esposito (2000) found that older adults have difficulty with an increased test load, as evidenced by a performance decrement when considering multiple items at test (independent of age-related difficulties with relational memory judgments). In the combination condition of the present research, to reject a new combination at test, one likely considers the test picture as well as one or both of the original pictures from which it was composed. Faced with this increased test load, younger children may have a greater tendency than older children for responding on the basis of familiarity (e.g., Jones & Atchley, 2002; Jones & Jacoby, 2001; Jones, Jacoby, & Gellis, 2001), even though they may have the ability to recollect the original stimuli. The finding that significant correlations between relational memory and episodic memory emerged only for the delayed test of the combination condition is consistent with a retrieval account: In effect, better retrieval processes may have led to better memory for the relations in the pictures and also to better free recall, particularly after an extended delay.

In fact, it is likely that there are two distinct developmental processes that the current data do not allow us to assess independently of each other: (a) early childhood improvement in the encoding of co-occurrence relations and (b) simultaneous improvement in retrieval of such relations. Supporting this dual hypothesis is the literature on neural development, which converges on the claim that prefrontal cortical regions—areas known to be involved in both episodic encoding and retrieval (for discussions, see Wheeler, Stuss, & Tulving, 1995, 1997)—undergo slow development relative to other cortical regions. For example, Huttenlocher and Dabholkar (1997) found in humans that synapse elimination of the middle frontal gyrus continues into mid-adolescence, compared with earlier termination of this process in the auditory cortex (see also Huttenlocher, 1979, 1990). Future research will need to address the different contributions of encoding and retrieval processes to early childhood memory development, providing a richer understanding of how children eventually begin to remember the past in an adultlike fashion.

References

Allen, B. P., & Lindsay, D. S. (1998). Amalgamations of memories: Intrusion of information from one event into reports of another. *Applied Cognitive Psychology, 12,* 277–285.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5,* 325–337.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24,* 403–416.

Chalfonte, B. L., Verfaellie, M., Johnson, M. K., & Reiss, L. (1996). Spatial location memory in amnesia: Binding item and location information under incidental and intentional encoding conditions. *Memory, 4,* 591–614.

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system.* Cambridge, MA: MIT Press.

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

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.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40,* 177–190.

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

Hannigan, S. L., & Reinitz, M. T. (2000). Influences of temporal factors on memory conjunction errors. *Applied Cognitive Psychology, 14,* 309–321.

Hannigan, S. L., & Reinitz, M. T. (2003). Migration of objects and inferences across episodes. *Memory & Cognition, 31,* 434–444.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging, 4,* 106–112.

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

Hashtroudi, S., Johnson, M. K., Vnek, N., & Ferguson, S. A. (1994). Aging and the effects of affective and factual focus on source monitoring and recall. *Psychology and Aging, 9,* 160–170.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33,* 1–18.

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

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.

Johnson, M. K., & Chalfonte, B. L. (1994). Binding complex memories: The role of reactivation and the hippocampus. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 311–350). Cambridge, MA: MIT Press.

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

Jones, T. C., & Atchley, P. (2002). Conjunction error rates on a continuous recognition memory test: Little evidence for recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 374–379.

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45,* 82–102.

Jones, T. C., Jacoby, L. L., & Gellis, L. A. (2001). Cross-modal feature and conjunction errors in recognition memory. *Journal of Memory and Language, 44,* 131–152.

Kirasic, K. C., Siegel, A. W., & Allen, G. L. (1980). Developmental changes in recognition-in-context memory. *Child Development, 51,* 302–305.

Kroll, N. E. A., Knight, R. T., Metcalfe, J., Wolf, E. S., & Tulving, E.

(1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language, 35,* 176–196.

Lindsay, D. S., Allen, B. P., Chan, J. C. K., & Dahl, L. C. (2004). Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another event. *Journal of Memory and Language, 50,* 96–111.

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

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

Markham, R. (1991). Development of reality monitoring for performed and imagined actions. *Perceptual and Motor Skills, 72,* 1347–1354.

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.

Moscovitch, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 269–310). Cambridge, MA: MIT Press.

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.

Pillemer, D. B., & White, S. H. (1989). Childhood events recalled by children and adults. *Advances in Childhood Development and Behavior, 21,* 297–341.

Reinitz, M. T., & Hannigan, S. (2001). Effects of simultaneous stimulus presentation and attention switching on memory conjunction errors. *Journal of Memory and Language, 44,* 206–219.

Reinitz, M. T., & Hannigan, S. L. (2004). False memories for compound words: Role of working memory. *Memory & Cognition, 32,* 463–473.

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition, 20,* 1–11.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 161–168.

Reinitz, M. T., Verfaellie, M., & Milberg, W. P. (1996). Memory conjunction errors in normal and amnesic subjects. *Journal of Memory and Language, 35,* 286–299.

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.

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

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49,* 289–318.

Shastri, L. (2002). Episodic memory and cortico-hippocampal interactions. *Trends in Cognitive Sciences, 6,* 162–168.

Sluzenski, J., Newcombe, N., & Ottinger, W. (2004). Changes in reality monitoring and episodic memory in early childhood. *Developmental Science, 7,* 225–245.

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.

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

Wechsler, D. (1989). *Manual for the Preschool and Primary Scales of Intelligence—Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

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

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,* 331–554.

Appendix

Story Used in the Free-Recall Task in Experiments 2 and 3

Wanda got out of bed on Saturday morning and saw that the weather was very rainy. She decided to bake a chocolate cake. So she went into the kitchen, and found the flour, eggs, sugar and chocolate that she would need to bake the cake. Wanda mixed everything together, poured the mixture into a large pan and put the pan into the oven. It was supposed to bake for an hour. While the cake was baking, the rain stopped and the sun came out. So Wanda put on her sneakers and went outside to play. Outside she found her friend Tyler. They played soccer together for a long time until they were tired. Then Wanda and Tyler decided to get some ice cream so they walked to the ice cream shop. They both bought vanilla cones and then began walking home. Suddenly, Wanda remembered the cake. She ran home, opened the oven, and found that the cake was burned! She had left it in for much too long.

Received January 4, 2005
Revision received August 22, 2005
Accepted August 27, 2005 ■