CHILD DEVELOPMENT

Child Development, September/October 2009, Volume 80, Number 5, Pages 1321–1328

Memory Binding in Early Childhood: Evidence for a Retrieval Deficit

Marianne E. Lloyd
Seton Hall University

Ayzit O. Doydum
Emory University

Nora S. Newcombe
Temple University

Previous research has suggested that performance for items requiring memory-binding processes improves between ages 4 and 6 (J. Sluzenski, N. Newcombe, & S. L. Kovacs, 2006). The present study suggests that much of this improvement is due to retrieval, as opposed to encoding, deficits for 4-year-olds. Four- and 6-year-old children (N = 48 per age) were given objects, backgrounds, and object + background combinations to remember. Younger children performed equivalently to 6-year-olds during a working memory task for all types of memory questions but were impaired during a long-term memory task for the object + background combinations. Furthermore, this deficit was completely due to differences in false alarm rates, suggesting that separate analyses of hits and false alarms may be preferable to corrected recognition scores when studying memory development.

Successful recognition memory performance requires the ability to discriminate the familiar from the novel (e.g., Have I ever seen this cat before?). Additionally, however, we often need to discriminate among familiar elements in terms of their prior co-occurrences (e.g., Have I ever seen this cat in this yard before?). That is, memory does not consist of single items but rather of complex combinations of items. The memory process involved in establishing connections between items (e.g., the cat and the yard), within items (e.g., that a particular car had red paint, a spoiler, and a leather interior), or between items and their contexts has been termed *memory binding* and is thought to be critical for episodic (e.g., Eichenbaum, 1997) and autobiographical (Newcombe, Lloyd, & Ratliff, 2007) memory. Successful memory-binding processes are crucial for later retrieval of memories, such as an accurate report on the joys of a child's graduation or upon the heartbreak of a very close basketball game.

There is a good deal of work on memory changes during aging that involves examination of memory-binding processes (e.g., Chalfonte & Johnson, 1996; Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000). In these studies, participants are given items to remember that vary in identity and a second feature such as color or location. For example, they might see a pink line drawing of an ironing board in the lower left corner of a screen and a green line drawing of a dog in the middle of the page. Participants then receive a memory test that requires them to remember single features (e.g., something pink? an ironing board? something in the lower left?) or the combination of features (e.g., a green ironing board?). The general pattern of results is that older adults show good performance on individual features but impaired performance, relative to younger adults, on combinations of features (e.g., Chalfonte & Johnson, 1996; Mitchell et al., 2000). However, even when older adults show some impairment in memory for features, they also show significantly larger deficits for the combinations of features (Mitchell et al., 2000). These results suggest that older adults are impaired in remembering the relations between items rather than limited in the number of individual items that can be remembered (for a similar idea, see Naveh-Benjamin, 2000). However, it is not the case that older adults always perform more poorly. One factor that determines the magnitude of the deficit seems to be the length

We would like to thank Hannah Bingman and Wendy Shallcross for their assistance in testing participants, and Stacie Crawley, Julie Sluzenski, and Laura Sywulak for their conversations pertaining to project development.

Correspondence concerning this article should be addressed to Marianne E. Lloyd, Department of Psychology, Seton Hall University, 400 South Orange Ave., South Orange, NJ 07079. Electronic mail may be sent to lloydmar@shu.edu.

of time between encoding and retrieval. When the delay is very short (< 2 s), older and younger adults perform similarly (e.g., Olson et al., 2004).

In terms of memory development in children, one might predict that successful recognition of feature combinations may develop later than processes that support memory for individual features. Autobiographical memory development can be conceptualized as a shift from early reliance on semantic memory to the later addition of competency in episodic memory due to developments in the hippocampus (e.g., O'Keefe & Nadel, 1978) and/or prefrontal cortex (e.g., Conway, Pleydell-Pearce, Whitecross, & Sharpe, 2003). In a memory task requiring successful memory-binding processes, intact and rearranged items share features that were both recently seen during the study phase. To discriminate intact from rearranged items, a child must also appreciate that how items were put together spatially and temporally (critical elements of Tulving's, 1983, definition of episodic memory) is important for memory decisions and must retain the links.

To date, there has been very little published work directly examining memory-binding processes in younger children, although two studies have looked at binding tasks with school-age children (Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Lorsbach & Reimer, 2005). In a recent study that was focused on memory performance in preschool-aged children, Sluzenski et al. (2006) investigated memory binding across early childhood by looking at interitem binding of animals and their backgrounds. Children studied an animal as well as its associated background and then were shown the animal and background paired together (Sluzenski et al., 2006). For example, a child might be shown a picture of a tiger, then a picture of a swimming pool, and finally a picture of the tiger at the swimming pool. In addition, she might also see a turtle, a library, and a picture of a turtle in a library. Later, the participants were given a memory test in which they had to discriminate studied animals (e.g., a tiger) from novel animals (e.g., a seal), studied backgrounds (e.g., a swimming pool) from novel backgrounds (e.g., a freeway), and in the object + background task, correct pairings (tiger in a swimming pool) from rearranged pairings (tiger in a library). The children performed equally well on the animal and background memory tasks but 4-year-olds performed more poorly on the tasks requiring discrimination of original and rearranged animals and backgrounds, suggesting a deficit in memory-binding processes similar to that reported in older adults (e.g., Chalfonte & Johnson, 1996; Mitchell et al., 2000).

Sluzenski et al. (2006) comment that reasons for the reduction in performance on test items thought to require memory-binding processes relative to feature memory is uncertain (see also Lorsbach & Reimer, 2005). It is possible that younger children simply encode less information and, specifically, do not encode the linkage between items and context. On the other hand, it is possible that they have difficulty with retrieval of the linkages. The goal of this study was to attempt to specify the nature of the deficit in memory performance for combinations of features by manipulating the time between encoding and retrieval. If the deficit is at encoding, then memory tasks requiring binding processes should still show age-related differences. However, if the age difference is due to poor retrieval, then performance should improve at shorter study–test intervals. Specifically, we expected to replicate—using new stimuli—the prior demonstrations of a deficit in long-term memory (LTM) tasks dependent on binding processes (e.g., Sluzenski et al., 2006) and determine whether performance would also be impaired for 4-year-olds in a working-memory (WM) test.

Method

Participants

Participants were recruited from the suburbs of Philadelphia. Most participants were Caucasian and from homes of a middle- to upper-class socioeconomic status. Two 6-year-old (1 in WM and 1 in LTM) and eight 4-year-old (5 in WM and 3 in LTM) participants were discarded for failure to follow the instructions during the study (e.g., failing to attend to the pictures during the encoding phase or requesting to terminate their participation). Four participants (three 4-year-olds and one 6-year-old) were removed due to their memory performance being a statistical outlier in the object false alarm condition; they were the only four children to make any object false alarms in the LTM group. This left a final sample of 96 participants. There were forty-eight 4-year-olds (M = 53.4 months, N = 25 females) and forty-eight 6-year-olds (M = 76.9 months, N = 26 females).

Design

For both the WM and LTM conditions, the experimental design was a 3 (test item type: object, background, and object + background) × 2 (test item status: old or new) × 2 (age: 4 or 6 years). Test item type and test item status were manipulated within participants. Age was manipulated between participants.

Apparatus and Materials

The training and tests were administered on a laptop computer with a 15-in. LCD screen accompanied by verbal instructions and questions posed by the experimenter. The experiment was presented and slides were timed using Microsoft PowerPoint 2003. The PowerPoint backgrounds were downloaded from the Microsoft Office Web site and were selected for distinctness from each other as well as minimal complexity so as not to overwhelm the object images and compete for attention. Unlike the stimuli from Sluzenski et al. (2006), which contained complex backgrounds made of several distinct features (e.g., buildings), the present backgrounds were chosen to be akin to a single feature. The objects were simple line drawings from the Snodgrass and Vanderwart norms (1980).

Procedures

Training phase. To maximize understanding of the task, a training procedure similar to the later test procedure allowed participants to become familiar with the question types and permitted them to revisit the stimuli in case they had trouble recalling the images. The training procedure first gave the participants separate practice with the three types of study items (objects, backgrounds, and objects + backgrounds). Specifically, participants were trained on two sets of study–test phases for each of the three types of study slides (objects, backgrounds, and objects + backgrounds). For these training trials, two encoding items were presented and then a test slide. After these six study–test phases, four study–test phases with mixed encoding items were presented. During these training trials, four encoding items that were mixed among object, background, and object + background study slides were presented followed by two test slides. For all encoding trials, participants were instructed to pay attention to the thing (the object) that they saw and/or the background of the picture presented on the computer screen. During the study phase, the experimenter always labeled all of the objects to eliminate the effect of some children's voluntary labeling. For slides when there was a background and an object present, the experimenter indicated with her hand the surrounding area around the object and instructed the participant that the background was ''where the thing belonged.''

During the recognition memory test, when an object + background test slide was presented, the participant was asked, ''Does this thing belong in this background, like when I just showed it to you?'' When a background test slide was presented, the participant was asked, ''Did you see this background before?'' When an object test slide was presented, the participant was asked, ''Did you see this picture before?'' Participants were instructed to answer ''yes'' or ''no'' to each test question.

Background questions consisted of a background from one of the encoding stimuli in the current trial or a completely novel background. Similarly, object questions consisted of an object that was either completely novel or taken from one of the slides in the current trial. In contrast, object + background lure questions consisted of an object paired with a background from a different studied object + background slide. By using two familiar features that had been re-paired, participants were prevented from responding correctly on the basis of discrimination between old and new features. For instance, if two of the encoding stimuli consisted of a cat on a green background and a cup on a red background, then an object + background lure question would be either the image of a cat on a red background or a cup on a green background. All participants viewed the same training presentation.

Study and testing phase. Participants were randomly assigned to either the WM or the LTM condition. During the study phase, participants were presented with a random assortment of encoding stimuli. Again, these consisted of a mixture of individual object images in the center of the screen, backgrounds, or both object images and backgrounds paired together. The presentation was followed by several test questions that were either feature based (object or background) or memory binding based (object and background combination). A sample encoding and test phase for both the LTM and WM tasks is depicted in Figure 1. The type of encoding stimuli dictated what type of test item could be created for the item. Thus, when participants studied a picture of only a bear, the test slide could be either a picture of a bear or a picture of an object that was not studied. That is, single backgrounds and single objects were always used for the object and background test questions. The object + background slides were only ever tested as object + background (either identical or re-paired)

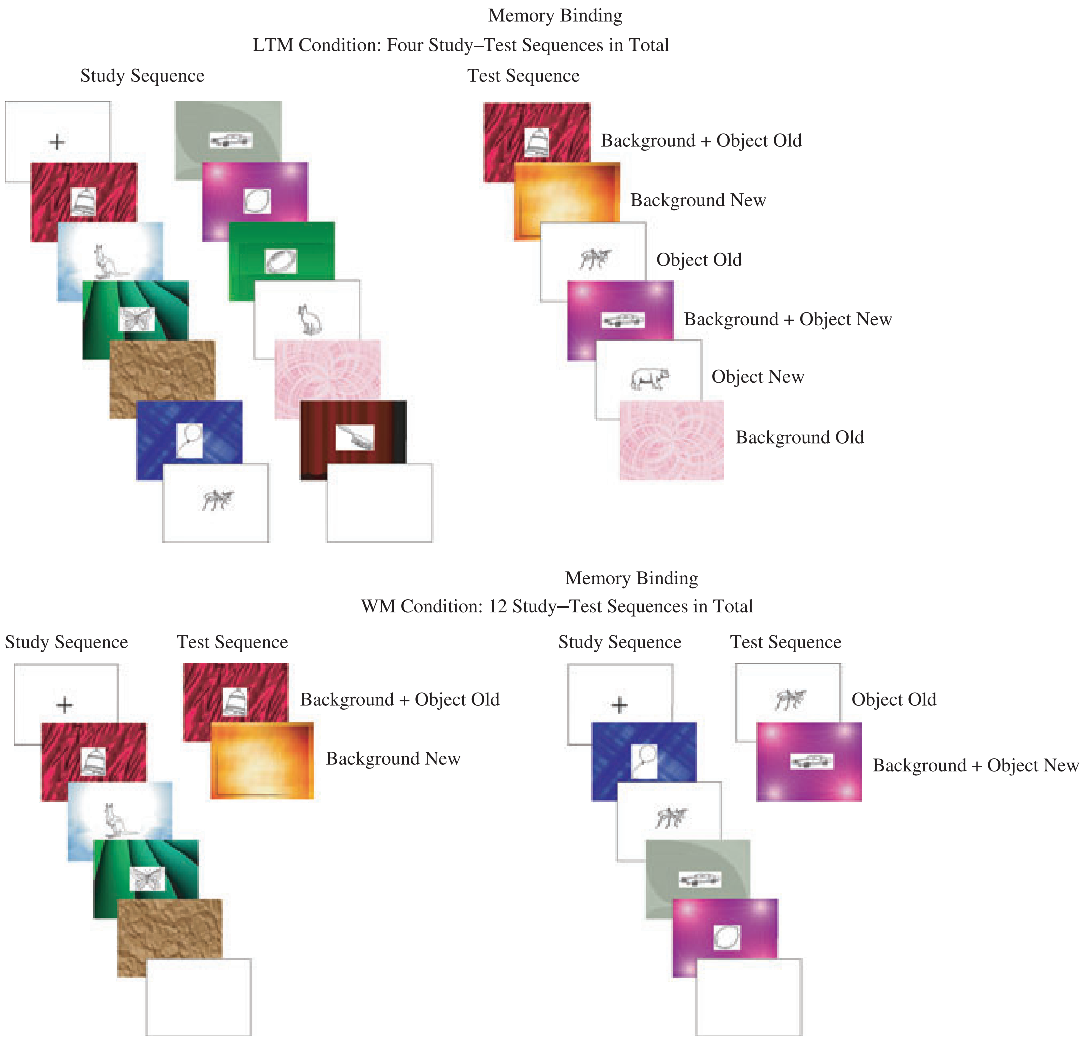

Figure 1. Study and test sequence for long-term memory (LTM) and working-memory (WM) condition. The major difference between the two conditions was the number of study items that were presented before the memory test. Participants either studied 12 (LTM) or 4 (WM) items and were then given a memory test.

items. It was never the case that an object that had been presented paired with a background was later tested as only an object. Participants were made aware of this rule during the training phase.

For the WM condition, the procedure consisted of 12 blocks in which each block was composed of four encoding stimuli presented sequentially, for 4 s each, followed by a 5-s delay period, and then the presentation of two test questions. For the LTM presentation, the procedure consisted of four test blocks in which each block consisted of 12 encoding stimuli presented for 4 s each, followed by a 5-s delay period, and then the presentation of six test questions. The delay periods in both presentations consisted of a plain white slide. These study–test blocks were repeated until each participant had studied 48 stimuli and answered 24 memory questions (8 object, 8 background, and 8 object + background questions). Half of the test items were old (studied object, studied background, or intact object + background) and half of the test items were new (novel object, novel background, rearranged object + background). The order of encoding stimuli and test items was counterbalanced across participants. Items served equally often as targets or lures. The relation of an item in the study phase to its position in the test phase was held constant. For WM tests, the first test question involved one of the first two encoding slides (or was an unrelated lure) and the second test item was generated from the final two encoding items. Similarly, for LTM tests, the six test questions were generated in order from the 12 encoding slides. Thus, an item from the LTM condition never had a delay between encoding and test that was similar to that of the WM group. There were an equal number of target and lure items for each question type (object, background, and object + background).

To summarize, all participants started with a training phase to teach them about the three types of study and test items (object, background and object + background). They were then given a series of study and test sequences. These study–test phases were repeated until 48 encoding trials and their associated recognition test questions had been administered. The entire procedure took approximately 30 min.

Results and Discussion

A significance criterion of .05 was used for all analyses reported here. Figure 2 shows mean ''yes'' responses for the full data set. Unlike Sluzenski et al. (2006), the present analyses were conducted for hit and false alarm rates separately rather than for corrected recognition—in which the false alarm rate is subtracted from the hit rate. If younger children have a specific memory binding deficit, then the false alarm rate should be higher than that of 6-year-olds, but the hit rate could be similar if the participants are basing their responses on the overall familiarity of the test stimuli.

As may be seen in Figure 2, the LTM results replicate those of Sluzenski et al. (2006). A series of planned comparisons were conducted to compare each of the six test conditions as a function of age. Feature memory was equivalent between age groups for background hits and false alarms (all *p*s > .15). The hit rate for objects was slightly higher for 6-year-olds (M = 1.00) than for 4-year-olds (M = .92), $t(46) = 2.89$, $p < .05$. The false alarm rate

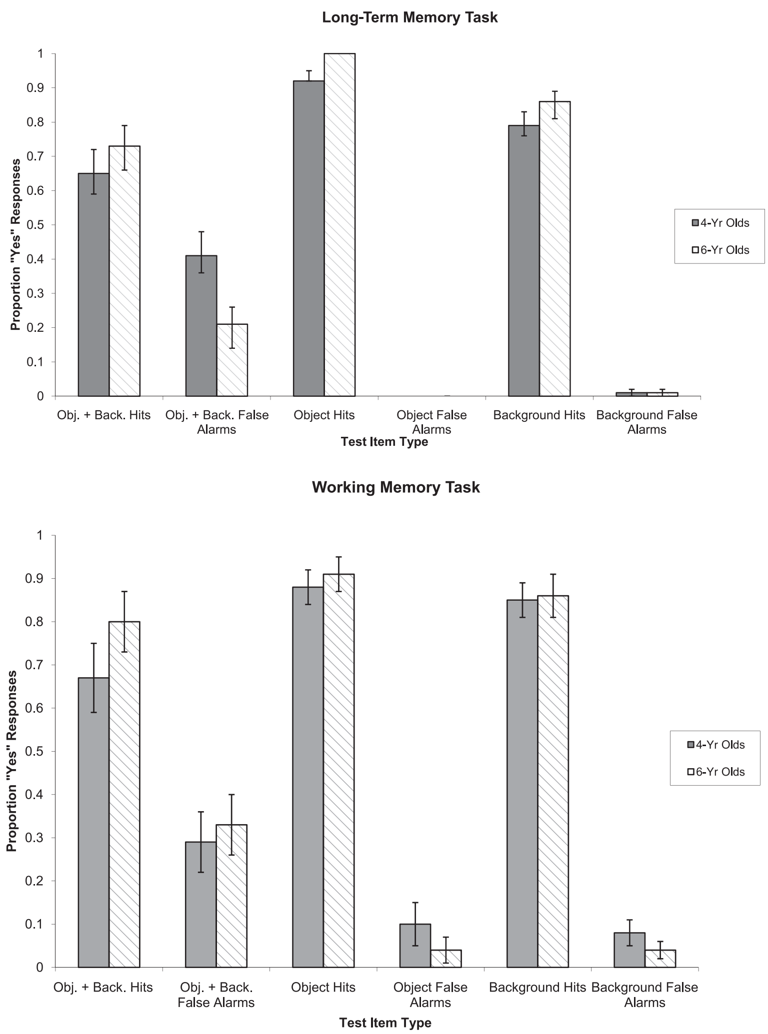

Figure 2. Mean ''yes'' responses as a function of age, test item type, question type, and delay condition.

did not differ because no participant made a false alarm to a novel object. In terms of the object + background test items, the hit rates were equivalent ($p < .40$) but 4-year-olds made significantly more false alarms than 6-year-olds, $t(46) = 2.34$. This suggests that it is not the case that the 4-year-olds had more difficulty on items that contained two features, but rather they had difficulty with remembering the way in which the objects were initially presented with backgrounds—that is, the binding or linkage of the items.

As was noted earlier, 6-year-olds in the LTM group had a significantly higher hit rate to objects in comparison to the 4-year-olds. By selecting a subset of data for which all of the 4-year-olds also had a perfect hit rate in the object condition, we attempted to address the possibility that reduced feature memory led to a decreased memory performance in the object + background condition. Doing so removed seven 4-year-olds from the analysis but did not change the pattern of results. Four-year-olds still made more false alarms to combination items (38%) than 6-year-olds (21%) while maintaining an equivalent hit rate with older children for such items (68% and 73% for 4- and 6-year-olds, respectively). This suggests that the finding of a difference in memory for combinations of features is not simply explained by better overall performance by the 6-year-olds (e.g., Lorsbach & Reimer, 2005).

The results of the WM group did not follow the same pattern as the LTM group because, in contrast to the results for LTM, the false alarm rate for object + background items was equivalent across age, $p < .63$. This result of an equivalent false alarm rate is in contrast to the finding of more false alarms for object + background questions by the 4-year-olds in the LTM condition. As was found in the LTM group, the object + background hit rate was equivalent, $p < .21$. In addition, the hits and false alarm rates for both objects and backgrounds were equivalent (all ps > .28).

To summarize, the results of the study suggest that the decreased performance for object + background items in 4-year-olds is a product of an increased false alarm rate for a LTM test relative to a WM test. This suggests that the problem is at least partially attributable to failures at retrieval rather than a failure to encode the items that were presented. Such a conclusion has been impossible in prior studies due to a lack of comparability between the methods. However, what cannot be determined from the current study is the exact reason for this difference. That is, the LTM condition had a longer delay between study and test as well as a larger number of encoding items. Future research is necessary to determine whether list length or delay (or both) lead to enhanced false alarms for re-paired object + background items.

The results presented here place a greater focus on differences in false alarm rates than other published studies investigating binding memory processes in children, which have used a corrected recognition score. The present study suggests that corrected recognition may not be the optimal means of analyzing such data. Indeed, a corrected recognition analysis would suggest that the 4-year-olds were deficient in object + background questions, but would mask the finding that this was due solely to the increase in false alarms. In contrast, there were no age-related differences in false alarms for either of the features, suggesting that it is not simply the case that 4-year-olds have a higher overall error rate on recognition memory tests. In comparing the results of the other studies to those presented here, there is evidence for a similar pattern. Sluzenski et al. (2006) found 4- and 6-year-olds were equally good at recognizing the items that had occurred together initially (hit rates of .76 and .79, respectively) but 4-year-olds had greater difficulty than 6-year-olds in rejecting items that were rearranged (false alarm rates of .36 and .20, respectively). Similarly, the results of Lorsbach and Reimer (2005) showed much larger differences in the false alarm rate for questions requiring memory binding processes than in the feature false alarm rates.

One theoretical explanation for the different rates of false alarms for 4-year-olds in WM and LTM tasks is based on the difference between two processes thought to underlie recognition memory, familiarity, and recollection (for a review, see Yonelinas, 2002). Whereas false alarms to objects or backgrounds were made to novel items, false alarms to object + background items were made to previously presented features that had been re-paired. Thus, the rearranged items would be familiar in a global sense, and a child could avoid an erroneous endorsement that the item was studied only by recollecting that the object had been paired with another background. Reliance on recollection has been proposed as a way of explaining how adults avoid similar memory errors (Lloyd, 2007; Odegard & Lampinen, 2004; Odegard, Lampinen, & Toglia, 2005). There is evidence that familiarity develops sooner than recollection in children (Ghetti & Angelini, 2008). Thus, one plausible explanation for the present results is a difference in recollection accuracy for WM tasks relative to LTM tasks.

The development of familiarity and recollection is relevant to, and not incompatible with, the distinction between semantic and episodic memory development. Reder et al.'s (2000) theory of recognition memory suggests that familiarity is increased by the strengthening of semantic concepts whereas recollection is a function of episodic activation. Thus, early development of familiarity relative to recollection would be consistent with shifts from semantic to episodic memory proficiency (e.g., Newcombe et al., 2007).

Two broad conclusions may be drawn from the work presented here. First, previous reports of a deficit in memory tasks requiring binding processes for preschool children are likely based at least partly on retrieval deficits, a conclusion similar to that of early studies of children and paired associate learning (e.g., Ackerman & Emmerich, 1978; Kee, Bell, & Davis, 1981). When the delay between study and test and/or the number of items between study and test is sufficiently short—in this case < 1 min—4-year-old children perform just as well as 6-year-old children. This conclusion is in contrast to that with older adults (e.g., Chalfonte & Johnson, 1996; Mitchell et al., 2000) in which memory tasks requiring binding processes is impaired in both WM and LTM tasks (although not after *very* short encoding–test delays; Olson et al., 2004). The current study is not sufficient to conclude, however, that encoding is not an important consideration for memory performance on these tasks. For example, it may be that if encoding of the object + background relations is weaker for younger children, a short delay before test or a fewer number of intervening items, is still sufficient to support successful memory binding processes. With longer delays, a weak encoding of relations may lead to reduced performance on the memory test. These weakened memory traces may then be less likely to be successfully recollected during the memory test, given that delays between study and test (e.g., Knowlton & Squire, 1995) and an increase in study items (e.g., Yonelinas & Jacoby, 1994) decrease the probability of recollection in adults. Future research into the relative influence of retrieval versus encoding to memory decisions is necessary to determine exactly which of many possible factors lead to a decrease in memory performance by 4-year-olds relative to 6-year-olds.

The second conclusion is that the standard of corrected recognition scores may be inappropriate for some types of data depending on the theoretical processes underlying hits versus false alarms. In the present study, a child's rates of false alarms were uniquely able to dissociate response based on semantic and episodic memory. We encourage memory development researchers to consider that, similar to the adult literature, as much knowledge may be gained from memory errors as correct memory decisions.

References

Ackerman, B. P., & Emmerich, H. J. (1978). When recognition memory fails: The use of an elimination strategy by young children. *Developmental Psychology*, *14*, 286–293.

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

Conway, M. A., Pleydell-Pearce, C. W., Whitecross, S. E., & Sharpe, H. (2003). Neurophysiological correlates of memory for experienced and imagined events. *Neuropsychologia*, *41*, 334–340.

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology*, *42*, 1089–1102.

Eichenbaum, H. (1997). How does the brain organize memories? *Science*, *277*, 330–332.

Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development*, *79*, 339–358.

Kee, D. W., Bell, T. S., & Davis, B. R. (1981). Developmental changes in the effects of presentation mode on the storage and retrieval of noun pairs in children's recognition memory. *Child Development*, *52*, 268–279.

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 699–710.

Lloyd, M. E. (2007). Metamemorial influences in recognition memory: Pictorial encoding reduces conjunction errors. *Memory & Cognition*, *35*, 1067–1073.

Lorsbach, T. C., & Reimer, J. F. (2005). Feature binding in children and young adults. *Journal of Genetic Psychology*, *166*, 313–327.

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

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1170–1187.

Newcombe, N. S., Lloyd, M. E., & Ratliff, K. R. (2007). Development of episodic and autobiographical memory: A cognitive neuroscience perspective. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 35, pp. 37–85). San Diego, CA: Elsevier.

Odegard, T. N., & Lampinen, J. M. (2004). Memory conjunction errors for autobiographical events: More than just familiarity. *Memory*, *12*, 288–300.

Odegard, T. N., Lampinen, J. M., & Toglia, M. P. (2005). Meaning's moderating effect on recollection rejection. *Journal of Memory & Language*, *53*, 416–429.

O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a cognitive map*. Oxford, UK: Clarendon Press.

Olson, I. R., Zhang, J. X., Mitchell, K. J., Johnson, M. K., Bloise, S. M., & Higgins, J. A. (2004). Preserved spatial memory over brief intervals in older adults. *Psychology and Aging*, *19*, 310–317.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Avers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 294–320.

Sluzenski, J., Newcombe, N., & Kovacs, S. L. (2006). Binding, relational memory, and recall of naturalistic events: A developmental perspective. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *32*, 89–100.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Tulving, E. (1983). *Elements of episodic memory*. Oxford, UK: Oxford University Press.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, *48*, 516–535.