Published in final edited form as:
J Exp Child Psychol. 2013 February ; 114(2): 321–338. doi:10.1016/j.jecp.2012.10.004.

Beyond initial encoding: Measures of the post-encoding status of memory traces predict long-term recall in infancy

Thanujeni Pathman and
Department of Psychology, Emory University, and Center for Mind and Brain, UC Davis

Patricia J. Bauer
Department of Psychology, Emory University

Abstract

The first years of life are witness to rapid changes in long-term recall ability. In the present research, we contributed to explanation of the changes by testing the absolute and relative contributions to long-term recall of encoding and post-encoding processes. Using elicited imitation, we sampled the status of 16-, 20-, and 24-month-old infants' memory representations at various time points after experience of events. In Experiment 1, infants were tested immediately, 1 week after encoding, and again after 1 month. The measure of 1-week trace status was a unique predictor of 1-month delayed recall. In Experiment 2, infants were tested immediately, 15 minutes, 48 hours, and 2 weeks after encoding, and again 1 month later. The measures of 15-minute and 48-hour trace strength contributed unique variance in 1-month delayed recall. The findings highlight the need to consider post-encoding processes in explanations of variability in long-term memory in infancy.

Keywords

consolidation; elicited imitation; infants; recall; encoding; deferred imitation

The ability to recall past events is a basic human capacity on which we depend. It allows us to remember the names and faces of people we have met, places we have gone, and in essence is a central part of what makes us who we are. Memory for events from the past begins to develop early in life and undergoes pronounced changes throughout infancy and beyond (for reviews see Bauer, 2007; Hayne, 2004; Howe & Courage, 1993; Rose, Feldman & Janowski, 2004). For example, there are well documented developmental changes in the amount and how long children remember (see Bauer, 2007; in press; for reviews). However, the memory processes that contribute to these patterns of remembering and forgetting are less well explicated, especially in infancy. The mnemonic processes themselves are known: encoding, consolidation, and retrieval. Less well understood are the absolute and relative contributions that each of these processes make to the robustness of a memory representation over a delay. The objective of the present investigation was to examine the variance in long-term recall in the second year of life explained by measures of the encoding

Correspondence may be addressed to Thanujeni Pathman, Center for Mind and Brain, UC Davis, 267 Cousteau Place, Davis, California, 95618, USA; tpathman@ucdavis.edu.

and post-encoding status of memory traces with the goal of informing the determinants of remembering and forgetting in infancy.

For a memory to be available for retrieval, it first must be encoded (initial registration of information) and then consolidated (stabilized and integrated into long-term stores). Substantial research has made clear that there are age-related and individual differences in encoding. For example, in infant preferential looking paradigms, the number of seconds of familiarization required for a stimulus to be encoded (as evidenced by a novelty response) changes with age, with younger infants requiring more encoding time to produce a novelty response, relative to older infants (Rose, Gottfried, Melloy-Carminar, & Bridger, 1982). In addition, younger infants require more trials to learn multi-step event sequences to a criterion (learning to a criterion indicates that the material was fully encoded) compared to older infants (Howe & Courage, 1997). Indeed, across development, older children show evidence of more rapid encoding relative to younger children (Howe & Brainerd, 1989).

The second phase in the life of a memory, consolidation, has been relatively neglected in the infant memory literature. Originally hypothesized by Müller and Pilzecker (1900), consolidation is a post-encoding process by which initially labile memory traces are stabilized and integrated into long-term storage (see McGaugh, 2000; and Wixted, 2004; for reviews). It is thought to be subserved by medial-temporal structures (the hippocampus) in concert with cortex (e.g., Zola & Squire, 2000). Experimental evidence that memory traces undergo changes post encoding come from studies with animal models in which consolidation processes have been deliberately disrupted (e.g., see Eichenbaum & Cohen, 2001; Squire & Alvarez, 1995; for reviews). For example, animals are trained to encode a new association such as between a tone and an electrical shock, after which they experience a lesion of the hippocampus. When animals are lesioned shortly after learning (e.g., 7 days), their memory for the association is severely impaired, suggesting that the recently-formed memory trace was still undergoing hippocampally-dependent post-encoding processing. In contrast, when animals are lesioned after longer intervals post encoding (e.g., 28 days), memory is unimpaired, suggesting that post-encoding processes were complete (e.g., Kim & Fanselow, 1992; Takehara, Kawahara, & Kirino, 2003). These and other similar observations imply that for new memories to be effectively stored, they must undergo additional processing after encoding.

Though there is strong evidence that memory traces undergo additional processing post encoding, there are few studies in which the implications of the processing for long-term recall have been investigated in humans (see Wixted, 2004, for a review). In a study with adults, Bosshardt and colleagues (2005) measured retrieval-related brain activity (with BOLD fMRI) 10 minutes and 24 hours after encoding of word pairs. They found differences in hippocampal activity 24 hours relative to 10 minutes post encoding. As noted by the authors, the finding is consistent with animal models which suggest changes in hippocampal synaptic connections within 24 hours after encoding (e.g, Dudai & Morris, 2000). Moreover, the change in memory trace status had implications for recall, as indicated by correlations between hippocampal activation and retrieval success: the correlation was stronger after 24 hours than after 10 minutes. Bosshardt et al. (2005) thus provides evidence of post-encoding changes in memory that have implications for long-term recall. It also suggests a potentially productive means of examining the contributions of post-encoding processing to long-term recall in infants, namely, probing the status of the memory trace at different points in time post encoding and determining the variance in long-term recall explained by each test.

In the infant literature there are relatively few studies that allow examination of post-encoding processes as a potential source of age-related or individual variability in long-term recall. A test of the question requires a measure of the newly encoded trace and a measure of

the strength of the memory trace thereafter, during the period in which consolidation processes are thought to occur. Many studies of infant memory lack both of these features (e.g., Barr, Dowden, & Hayne, 1996). The few studies in the infant literature that permit a test of the variance in long-term recall explained by post-encoding processing indicate that it is an important source of variance. The evidence comes from studies using elicited imitation. In elicited imitation, props are used to produce an action or a sequence of action that the infant then is invited to imitate. As reviewed in detail elsewhere (e.g., Bauer, 2004, 2006, in press), there is substantial evidence that this type of imitation-based paradigm serves as a non-verbal analogue to verbal report, including evidence that it is supported by medial-temporal lobe structures (Adlam, Vargha-Khadem, Mishkin, & de Haan, 2005; McDonough, Mandler, McKee, & Squire, 1995). Using an imitation-based paradigm, Bauer, Wiebe, Carver, Waters, and Nelson (2003) exposed 9-month-old infants to event sequences and then tested for immediate and 1-week delayed recognition of the props used to produce the sequences. The immediate recognition test served as the measure of encoding trace status; the 1-week delayed recognition test served as the assessment of the status of the memory trace post encoding. One month later, the infants were tested for recall of the sequences. The variance in long-term recall associated with immediate recognition was nonsignificant, whereas the 1-week delayed recognition test explained 28% of the variance in long-term recall.

A similar pattern was obtained with infants in the second year of life. Bauer, Cheatham, Cary, and Van Abbema (2002, Experiment 4), assessed 20-month-olds' recall of multi-step event sequences immediately after seeing the sequences modeled (assessment of encoding trace status), again after 48 hours (assessment of memory trace status post encoding), and 1 month (measure of long-term recall). The measure of encoding trace status explained a nonsignificant amount of variance in long-term recall. In contrast, the measure of the status of the memory trace after 48 hours explained 25% of the variance in recall after 1 month. Other studies in the infancy literature also implicate post-encoding processes in the explanation for patterns of long-term recall: In the face of equal encoding, over a delay, infants evidence differences in memory as a function of different stimulus types (Bauer, Güler, Starr, & Pathman, 2011), and age (Bauer, 2005; Howe & Courage, 1997), that are not explained by differences in retrieval processes. In summary, the few studies that allow test of the implications of post-encoding processes suggest that they occur and are a significant source of variance in long-term recall.

The present research is a complement to and extension of the small literature that considers the role of post-encoding processes in explaining long-term recall in infancy. In two experiments, we tested the unique and combined variance in 1-month delayed recall explained by measures of the success of encoding and of post-encoding processes. In both experiments, participants were infants 16, 20, and 24 months of age. We focused on 16- to 24-month-old infants for two primary reasons. First, within the 16- to 24-month age range there is substantial age-related variability in memory behavior to be explained. In this brief period there is rapid developmental change in the amount of information infants remember (e.g., Burch, Schwade, & Bauer, 2010) and the length of time over which memory is apparent (e.g., Bauer, Wenner, Dropik, & Wewerka, 2000). Second, there is reason to expect that post-encoding changes in memory traces may play an important role in explaining the observed age-related variability. We make this conjecture in recognition of evidence that in this age range, the neural structures implicated in long-term recall are substantially immature (see Bauer, 2004, 2007, 2009, for discussions). That is, both the medial-temporal lobes and the association cortices involved in consolidation undergo post-natal developmental change throughout infancy and well beyond (e.g., Bauer, 2006, in press; Nelson, de Haan, & Thomas, 2006). It is logical to expect that as a result of the relative immaturity of the neural structures and network on which they depend, the processes of stabilization and integration

of memory traces will be relatively ineffective and thus a source of substantial variance. As a result, post-encoding processes are an especially attractive potential determinant of long-term recall in infancy (Bauer, 2006, 2007, 2008, 2009; Bauer et al., 2011); evidence consistent with this suggestion was provided in Bauer et al. (2002). Thus the period of 16 to 24 months is well suited to investigation of the variance in long-term recall explained by encoding and post-encoding processes.

In each of two experiments, we exposed 16- to 24-month-olds to multistep event sequences and then probed their memory for the sequences immediately after modeling, as a measure of encoding trace status. Also common to both experiments was a final test of long-term recall after 1 month. We chose this time frame because research with animal models suggests that by 28 days post encoding, memories are relatively stable (i.e., they are no longer dependent on the hippocampus, suggesting that at least a functional level of systems-level consolidation has occurred by that time: Kim & Fanselow, 1992; Takehara et al., 2003). As well, 1-month delays have been used in prior related studies (Bauer et al., 2002, 2003), thereby ensuring a direct point of comparison.

In between the measures of encoding and 1-month delayed recall were one (Experiment 1) and three (Experiment 2) measures of the status of memory traces in the post-encoding consolidation period. In Experiment 1, we probed memory trace status 1 week after encoding. To assess the absolute amount of variance in long-term recall explained by the 1-week measure of post-encoding trace status, for half of the event sequences, we brought the infants to a criterion level of learning, thus effectively eliminating variability in encoding as a source of variance in long-term recall. For the other half of the sequences, learning levels were free to vary. This manipulation permitted evaluation of the individual and combined contributions to long-term recall explained by the measure of the encoded trace and the measure of the status of the trace post encoding. In Experiment 2, we extended the approach by measuring the status of the memory trace at multiple times post encoding. Specifically, in addition to the immediate test of the encoded trace, we tested memory 15 minutes, 48 hours, and 2 weeks post encoding. Each test provided an independent assessment of the status of the memory trace at a point during the period of consolidation, thereby informing the course of functionally relevant post-encoding changes in trace status. The memory tests all were timed to occur during the period of active hippocampally-dependent post-encoding processing, as indicated by research with animal models (Kim & Fanselow, 1992; Takehara et al., 2003), which is virtually the only source of relevant data.

In summary, in the present research we examined 16-, 20-, and 24-month-old infants' recall of multi-step event sequences immediately after seeing them modeled, as a measure of what was encoded; at various times after encoding, as an assessment of post-encoding memory trace status; and 1 month later, as an assessment of long-term recall. The experiments provide strong tests of the absolute and relative contributions of measures of encoding and post-encoding processes for long-term recall in infants, and contribute to the relatively limited literature on the sources of variability in long-term recall in infancy.

Experiment 1

Method

Participants—Fifty-one infants participated: 17 16-month-olds (8 girls; mean age: 16.2 months; range: 15.6–16.5 months), 17 20-month-olds (8 girls; mean age: 20.0 months; range: 19.3–20.5 months), and 17 24-month-olds (9 girls; mean age: 24.1 months; range: 23.6–24.5 months). Infants were recruited from a departmental subject pool, comprised of names of parents whose infants were born in local hospitals and volunteered to participate in research. Three participants were African American and the remaining participants were

Caucasian. Infants received a small toy after each session, and parents received a $10 gift certificate to a local merchant after the last session. A university Institutional Review Board approved the protocol and procedures. All parents gave written informed consent for their infants' participation. An additional 6 infants participated but were excluded because they (a) did not return for their last visit (n=4), (b) did not return for their last visit within the specified delay (n=1), or (c) were born prior to 38 weeks gestation (n=1).

Materials—Stimuli were six multi-step sequences the orders of which were constrained by enabling relations. Enabling relations are said to exist when one step in a sequence is both temporally prior to and necessary for successful completion of the following step; enabling relations facilitate recall (e.g., Bauer, 1992; Bauer, et al., 2000). Infants were tested with different sequence lengths, determined appropriate for each age group based on prior research (Bauer & Dow, 1994; Bauer & Hertsgaard, 1993; Bauer & Travis, 1993; Burch et al., 2010). Two-step, three-step and four-step versions of sequences were used for 16-, 20- and 24-month-olds, respectively. The difference between the versions of the sequences was the number of steps necessary to reach the end-state or goal of the sequence (which was the same across step lengths). For example, the final step of each version of the sequence "make a gong" was to ring a metal plate with a mallet. The 4-step version of "make a gong" consisted of a base, a bar standing vertically in a cylinder attached to the base, a metal plate with a curved lip, and a wooden mallet. The steps of the sequence were lifting the vertical bar out of the cylinder (Step 1), placing the bar horizontally onto the two arms of the base (Step 2), hanging the plate on the bar (Step 3), and ringing the plate with a mallet (Step 4). For the 3-step version of this sequence, the first step from the 4-step version was not necessary, because the bar was hinged to one arm of the base, and thus the sequence required only folding the hinged bar to create the horizontal support (Step 1), hanging the plate on the bar (Step 2), and ringing the plate with a mallet (Step 3). In the 2-step version, the horizontal bar was attached to the base, and thus the only steps were hanging the plate on the bar (Step 1), and ringing the plate with a mallet (Step 2). All sequences were novel according to parent report, and had been used in previously published studies (e.g., Bauer et al., 2000, 2011). A list and description of the sequences is available from the authors.

Procedure—Participants took part in three sessions. Sessions 1 and 2 were separated by 1 week (M = 7.1 days; range 5–10 days); Sessions 2 and 3 were separated by 1 month (M = 28.2 days; range 25–35 days). The sessions took place in a laboratory testing room and were recorded on DVD. Each infant was tested by one of two experimenters; the experimenters tested an approximately equal number of infants at each age and gender. Infants were tested by the same experimenter at all sessions. Experimenters followed a detailed written protocol. Fidelity to the protocol was ensured by regular reviews of the DVDs of the sessions by the experimenters. For all sessions, the infant and the experimenter sat at a table across from each other (the infant sat in a booster seat), and the infant's parent sat beside the infant. Each session began with a 2–4 minute warm-up period during which the infant and experimenter played with commercially available toy beads. This warm-up period helped infants become comfortable with the experimenter and testing environment, and also helped infants understand the turn-taking nature of the elicited imitation paradigm.

Encoding session: After the warm-up period, infants were introduced to six sequences in turn. Half of the sequences were presented in a *standard* and half in a *criterion* encoding condition. In the standard encoding condition (three sequences), as in prior research using this paradigm (e.g., Bauer, 1992; Bauer et al., 2000), the experimenter provided the props for each sequence in turn and allowed the infant to interact with them for a baseline period. Baseline was "infant controlled" and ended when infants pushed away the stimuli, or engaged in repetitive exploratory behavior. After the baseline period, the experimenter used

the stimuli to model the actions of the sequence twice in succession. The experimenter labeled the steps in the sequence as she performed them. For example, for the 3-step "make a gong" sequence, the experimenter said: "This is how I make a gong. Put on the bar (putting the bar on the arms of the base). Hang up the bell (hanging the metal plate on the bar). Ring it (ringing the metal plate with the mallet). This is how I make a gong." After the second modeling, infants were given the opportunity to imitate. The experimenter placed the stimuli in front of the infant and said, "Now it is your turn! Show me how to [name of sequence], just like I did."

In the criterion encoding condition (three sequences), infants were given multiple opportunities to imitate until they showed evidence of maximum learning (i.e., complete encoding). After the first imitation period, the experimenter modeled the sequence again one time. The experimenter then returned the stimuli to the infant and encouraged imitation. This continued until the infant imitated the actions in target temporal order two times in a row (as in Bauer et al, 2011, and Howe & Courage, 1997), or until the infant was given a total of four imitation periods, whichever came first. A maximum of four imitation periods was imposed to avoid taxing or frustrating the infants. Across infants, sequences were used in the standard and criterion encoding conditions equally often. Within participants, the order of presentation of sequences alternated between standard and criterion. For half of the infants, the first sequence was in the standard condition and for half the infants, the first sequence was in the criterion condition.

Test of post-encoding trace status: One week later, infants returned for their second session in which they were tested on each of the six sequences in turn. For each sequence, the experimenter placed the props in front of the infant and said, "Show me what you can do with this stuff." The experimenter provided general encouragement (e.g., "Good job!") but no verbal labels or reminders of actions to be performed were given. At no point did the experimenter remodel the sequences. The order of presentation of the sequences was counterbalanced across infants and was different than at Session 1.

Long-term memory retrieval session: One month later, infants returned to the laboratory and were tested on all the event sequences. For each sequence in turn, the experimenter placed the props in front of the infant and said, "Show me what you can do with all this stuff." Again, the order of presentation of the sequences was counterbalanced across infants and was different than at Sessions 1 and 2.

Scoring: Each sequence was scored for the number of individual target actions produced (max = 2, 3, or 4, for 2-, 3-, and 4-step sequences, respectively), as a measure of item memory, and the number of pairs of actions produced in the target temporal order (max = 1, 2, or 3, for 2-, 3-, and 4-step sequences, respectively), as a measure of memory organization. In calculating the number of pairs of actions produced, only the first occurrence of each target action was considered. For example, in "make a gong," if an infant produced the actions 1-2-3, s/he received a score of 3 individual actions and 2 pairs of actions in target order (Step 1-Step 2, and Step 2-Step 3). However, if the infant produced actions 1, 3, 2, s/he received credit for 3 individual target actions and 1 pair of actions (Step 1-Step 3). If the infant produced steps 3, 1, 3, s/he received credit for 2 individual target actions and 0 pairs of actions (the second occurrence of action 3 was not considered). This scoring procedure reduces the likelihood of an infant receiving credit for the production of an ordered pair of actions by chance or trial and error. In addition, infants were given credit for individual target actions regardless of the order in which they produced them. For example, in "make a gong," credit was given for "ring it" regardless of whether the metal plate was hanging on the bar. Thus ordered production of actions was not dependent on the physical constraints

imposed by the stimuli. These scoring procedures have been used in several previous studies (see Bauer et al., 2000, 2011; Burch et al., 2010, for examples).

Prior to beginning testing, the experimenters trained together on an existing corpus of elicited-imitation data and established reliability at 90% or above on three successive participants. The experimenters then coded the behavioral performance of the infants online (i.e., during the session). For purposes of reliability, a third rater independently re-coded 25% of the participants (offline, from DVD). Overall agreement between the rater and experimenters (i.e., the number of agreed upon target behaviors divided by the total number of target behaviors recorded) on both occurrence and order of target behaviors was 92.3% (range = 88–96%).

Results

Though the same stimuli were used in the standard and criterion encoding conditions, the procedures were substantially different. For this reason, and because direct comparison of levels of performance under the different encoding conditions was not the purpose of the research, we do not report statistical comparisons between conditions. Rather, for the criterion and standard conditions in turn, we first present results of analyses to determine whether the infants learned the event sequences as a result of exposure to the model and remembered them over the delays. Because the infants in the different age groups were tested on sequences of different lengths, we conducted the analyses for each age group separately. We next report results of tests of the amount of variance in long-term recall explained by measures of post-encoding trace status alone (criterion encoding condition) and in combination with measures of the success of encoding (standard encoding condition). The analyses were conducted across age groups, to permit sufficient power. Dependent measures were rendered comparable by using proportion—rather than raw—scores.

Criterion Encoding Condition—FollowingBauer et al. (2011) and Howe and Courage (1997), the criterion for complete encoding was two successive accurate reproductions of the event sequences. Only data from participants who reached criterion on at least one sequence (of three) were included in analyses. Table 1 shows the number of participants in each age group who reached criterion on zero, one, two, or all three sequences. Across age groups, 11 participants were excluded from the analyses because they did not reach criterion on any of the sequences. For infants who reached criterion on only one sequence, scores were based on that event. For infants who reached criterion on two or all three sequences, scores were the average across the sequences on which the infants reached criterion.

Assessments of retention over 1 week and 1 month: Descriptive statistics for performance in the baseline phase, and on the 1-week and 1-month delayed recall tests are provided in Table 2, Panel a. Because, by definition, infants' performance at immediate imitation was perfect, scores for the immediate recall phase are not represented in the table. To determine whether the infants retained memories of the event sequences over 1 week and 1 month, we conducted repeated measures analyses of variance (ANOVAs) with three levels of test (baseline, 1-week, 1-month), for each age group and dependent variable separately. For all three age groups, there were main effects of test, for individual target actions produced (16-month-olds, $F(2,10)=24.7$, $p<.0001$, $\eta_p^2 = .71$; 20-month-olds, $F(2,14)=59.3$, $p<.0001$, $\eta_p^2 = .81$; 24-month-olds, $F(2,13)=120.1$, $p<.0001$, $\eta_p^2 = .90$) and for pairs of actions produced in target temporal order (16-month-olds, $F(2,10)=8.6$, $p<.005$, $\eta_p^2 = .46$; 20-month-olds, $F(2,14)=51.9$, $p<.0001$, $\eta_p^2 = .79$; 24-month-olds, $F(2,13)=93.5$, $p<.0001$, $\eta_p^2 = .88$). For each age group, pairwise comparisons (Bonferroni corrected, $p<.05$) revealed that for both individual target actions and pairs of actions in target order, baseline performance was lower than both 1-week and 1-month recall; 1-week and 1-month recall did not differ. Thus all

three age groups of infants showed evidence of memory for the sequences in the criterion encoding condition, and there was no evidence of forgetting between the 1-week and 1-month recall tests.

Explanation of variance in long-term recall: To determine whether the post-encoding measure of the status of the memory trace explained variance in long-term recall, we regressed 1-month recall on performance at 1-week recall. As shown in Table 3, Panel a, across ages (N=40), separate multiple linear regressions revealed that for the number of individual target actions produced and the number of pairs of actions produced in target order, 1-week recall predicted 76% and 61% of the variance in recall after 1-month (respectively). Thus with the variability associated with encoding experimentally removed, the post-encoding measure of memory trace strength was a strong predictor of long-term recall.

Standard Encoding Condition

Assessments of encoding and retention over 1 week and 1 month: Descriptive statistics for performance in the baseline phase and on the immediate, 1-week, and 1-month delayed recall tests are provided in Table 2, Panel b. To determine if there was learning and retention over 1 week and 1 month, we conducted repeated measures ANOVAs with four levels of test, for each age group and each dependent variable separately. For all three age groups, there was evidence of learning and retention over both delays, as evidenced by main effects of test and pairwise comparisons (Bonferroni corrected p<.05).

For 16-month-olds, there were main effects of test, for individual target actions, $F(3,16)=7.6$, $p<.0001$, $\eta_p^2 = .32$, and pairs of actions in target order, $F(3,16)=8.0$, $p<.0002$, $\eta_p^2 = .33$. Pairwise comparisons revealed that for both dependent variables, baseline performance was lower than immediate, 1-week, and 1-month recall. The measures of immediate, 1-week, and 1-month recall did not differ from one another. Thus 16-month-olds learned the sequences, and they showed no evidence of forgetting even after 1 month. For 20-month-olds, main effects of test were found for individual target actions, $F(3,16)=41.8$, $p<.0001$, $\eta_p^2 = .72$, and pairs of actions in target order, $F(3,16)=25.9$, $p<.0001$, $\eta_p^2 = .62$. Pairwise comparisons revealed that for both dependent variables, baseline performance was lower than immediate, 1-week, and 1-month recall. For both dependent variables, immediate performance was higher than 1-week and 1-month recall which did not differ from one another. Thus 20-month-olds learned the sequences and they remembered them over the delays; they experienced some forgetting by 1-week and no additional forgetting over 1 month. For 24-month-olds, main effects of test were found, for individual target actions, $F(3,16)=36.8$, $p<.0001$, $\eta_p^2 = .70$, and pairs of actions in target order, $F(3,16)=54.2$, $p<.0001$, $\eta_p^2 = .77$. Pairwise comparisons revealed that for both dependent variables, baseline performance was lower than immediate, 1-week, and 1-month recall. For individual actions, the immediate, 1-week, and 1-month recall tests did not differ. For pairs of actions in target order, immediate performance was higher than 1-month recall; 1-week performance did not differ from either immediate performance or 1-month recall. Thus 24-month-olds learned the sequences and they showed no evidence of forgetting of the individual actions of the sequences. Though there was no evidence of forgetting of the temporal order of the sequences over 1 week, 1-month delayed recall was lower than immediate recall.

Explanation of variance in long-term recall: To determine whether measures of encoding and post-encoding trace status were predictive of recall after 1 month, we regressed 1-month recall performance on the linear combination of performance at immediate recall and 1-week recall (Table 3, Panel b). Specifically, we entered the measure of encoding (immediate recall) into the model first (Step 1), and then we entered the measure of post-encoding trace

status (1-week recall) in the model (Step 2) to determine whether the measure of post-encoding trace status accounted for unique variance in long-term recall, over and above that explained by the measure of encoding. Within the model, beta weights were evaluated to determine whether immediate and/or 1-week recall were significant predictors of performance at 1-month.

Across ages (*N*=51), when it was the sole predictor in the model (Step 1), immediate recall was a significant predictor of performance after 1 month, explaining 52% and 53% of the variance in long-term recall of the individual actions and pairs of actions in target order, respectively. However when measures of both immediate recall and performance after 1 week were entered into the models (Step 2), whereas the models were significant, only the beta weights for 1-week performance were significant. The addition of 1-week performance into the models explained an additional 16% and 11% of the variance for individual target actions and pairs of actions in target order, respectively, bringing the total variance accounted for to 68% and 63%, respectively.

Discussion

The purpose of Experiment 1 was to examine the contributions of encoding and post-encoding processes to long-term recall of event sequences by infants 16 to 24 months old. We examined the question under conditions of complete encoding, ensured by requiring learning to criterion, and when the success of encoding was free to vary. When encoding was controlled, and thus the only predictor in the model, the measure of the strength of the memory trace 1 week post encoding predicted significant variance in recall after 1 month. This aspect of the experiment provided evidence that in the 16 to 24 month age range, post-encoding changes in the memory trace are a significant source of variance in long-term recall.

The pattern of findings in the standard encoding condition provided evidence that post-encoding changes in the memory trace are not only a source of variance in long-term recall, but that they explain more variance than measures of encoding. That is, when encoding was the only variable in the model, it predicted significant variance in long-term recall. However, when the measure of trace status after 1 week was entered into the model, it explained significant variance above and beyond that explained by the measure of encoding. Moreover, with the addition of 1-week recall into the model, encoding-related variance no longer was significant. The findings are consistent with other infant studies (e.g., Bauer et al., 2002, 2003) in which assessments during the period of consolidation of a memory trace have proven to be better predictors of long-term recall relative to tests of encoding trace status. Together, these studies imply that the success of post-encoding processes accounts for variance in long-term recall in infancy.

Experiment 1 is an important addition to the literature on the determinants of remembering and forgetting in infancy. Yet it is limited in the information that it can provide about the course of functionally relevant post-encoding changes in memory traces, because only one test was included between encoding and long-term retrieval. To more precisely determine the relevant parameters, in Experiment 2, in addition to the immediate test of the encoded trace and the 1-month delayed recall test, we tested memory 15 minutes, 48 hours, and 2 weeks post encoding. Each test was conducted on different event sequences, thereby permitting independent assessments of the memory traces at points during the period in which consolidation processes are presumed to occur. The 15-minute and 48-hour tests were included to allow assessments of the status of memory representations during the initial period of consolidation that is associated with changes in synaptic strength (e.g., Bosshardt et al., 2005). The 2-week test was included to provide assessment of the functional significance of later-stage changes in trace status that may be associated with changes in the

role of the hippocampus in maintaining the memory representation (e.g., Takehara et al., 2003).

The addition of a 2-week test in Experiment 2 also allowed us to evaluate an alternate explanation for the results of Experiment 1. Specifically, we suggest that the 1-week test was a strong predictor of long-term recall because it was sensitive to changes in the memory representation that take place shortly after encoding. An alternate explanation is that the 1-week test was simply closer to the 1-month test, relative to the measure of encoding, thereby rendering it a better approximation of the later memory representation. By this reasoning, the 2-week test should account for more variance in long-term recall, relative to measures obtained closer to the time of encoding. By including three post-encoding tests—two early in the period during which consolidation is thought to occur and one later in the period—we were able to evaluate these competing explanations for the pattern of findings.

Experiment 2

Method

Participants—Fifty-five infants participated. There were 18 16-month-olds (9 girls; mean age: 16.1 months; range: 15.6–16.4 months), 18 20-month-olds (9 girls; mean age: 20.1 months; range: 19.7–20.5 months), and 19 24-month-olds (11 girls; mean age: 24.1 months; range: 23.5–24.5 months). Participants were recruited from a similar population as in Experiment 1. Six of the infants were African American and the rest were Caucasian. Each infant received a small toy at the end of each session, and parents received a $10 gift certificate to a local merchant after the last session. University Institutional Review Boards approved the protocol and procedures, and all parents gave written informed consent for their infants' participation. An additional 7 infants were tested but not included in data analysis because they did not return for a session (n=6) or audio-visual recording error (n=1).

Materials—Stimuli were twelve 3-step event sequences, all constrained by enabling relations. Six sequences had been used in Experiment 1 and the others were drawn from previously published studies (e.g., Bauer et al., 2000; Wiebe & Bauer, 2005). All infants were tested with sequences of the same step-length for two reasons. First, it allowed all infants to be included in the same ANOVAs, and regression analyses could be conducted without conversion of scores to proportions. Second, we expected use of 3-step sequences to exploit individual variability within the group, while avoiding both floor and ceiling effects. Three-step sequences have been used without floor effects with infants 16 months of age (e.g., Bauer & Dow, 1994), and are sensitive to variation among infants as old as 24 months (Burch et al., 2010). All sequences were novel, according to parent report. Each infant was tested on eight of the 12 sequences. The event sequences were counterbalanced such that across infants, each sequence was used equally often and also was tested equally often in the different delay conditions.

Procedure—Participants took part in four sessions. Sessions 1 and 2 were separated by 2 days (M=2.0 days); Sessions 2 and 3 were separated by 2 weeks (M=12.8 days; range 10–17 days); Sessions 3 and 4 were separated by 1 month (M=28.7 days; range 24–36 days). Sessions took place in a laboratory testing room and were recorded on DVD. Each infant was tested by one of three experimenters. Infants were tested by the same experimenter at all sessions. As in Experiment 1, fidelity to the protocol was ensured by regular reviews of the session DVDs among the experimenters. Seating of infants, experimenters and parents, and procedure during the warm-up periods were similar to those described in Experiment 1.

Encoding session: After the warm-up, infants were introduced to event sequences. All sequences were presented in the standard encoding condition described in Experiment 1. As in Experiment 1, the experimenter modeled each sequence while labeling the event and narrating the steps involved, twice in succession. Infants then were given an opportunity to imitate. To minimize participant burden, we introduced two deviations from the procedure used in the standard encoding condition of Experiment 1. First, we did not include a baseline phase. Comparisons of baseline levels of performance with performance after modeling have been featured in numerous studies with infants in the 16 - to 24-month age range and in every case, performance after exposure to the modeled sequences has been significantly higher than performance in the baseline phase, indicating successful learning of the event sequences (e.g., Bauer & Dow, 1994; Bauer & Hertsgaaard, 1993; Bauer & Mandler, 1989; Bauer et al., 2000). Thus we are confident that infants' post-modeling performance can be attributed to their memory for the modeled sequences. Second, introduction of the event sequences (i.e., modeling and the immediate imitation test) was distributed across Session 1 and 2. As reflected in Table 4, Events A, C, D, E and G were introduced at Session 1, and events B, F, and H were introduced at Session 2. Although this feature of the design meant that the delay for the two events in the 2-week test condition differed by 48 hours (i.e., the delay for event F was 48 hours shorter than the delay for Event E), preliminary analyses (see below) revealed no difference between performance on the two sequences in this condition.

Tests of post-encoding trace status: Events A and B were tested 15 minutes after modeling (i.e., in the same session that they were modeled by the experimenter). Events C and D were tested 48 hours after modeling (at Session 2). Events E and F were tested 2 weeks after modeling (Session 3). The procedure was the same as followed in Experiment 1. That is, for each sequence in turn, the experimenter placed the props in front of the infant and said, "Show me what you can do with this stuff." The experimenter provided general encouragement but no verbal labels or reminders of actions to be performed were given. At no point were the sequences remodeled.

Long-term memory retrieval session: Events A, C, E, and G were tested at Session 4, 1 month after Session 3 (approximately 6 weeks after initial encoding). For each sequence, in turn, the experimenter placed the sequence props in front of the infant and said, "Show me what you can do with all this stuff."

Scoring: Each event sequence was scored for the number of individual target actions produced (max = 3) and the number of pairs of actions produced in the target temporal order (max = 2). The scoring procedure was the same as that used in Experiment 1. As in Experiment 1, experimenters coded the behavioral performance of the infants online. Prior to testing, experimenters were trained on scoring procedures using an existing corpus of data and established reliability at 90% or above on three successive participants. For purposes of reliability, two of the experimenters independently re-coded 25% of the third experimenter's participants (offline, from DVD). Overall agreement between the raters on both occurrence and order of target behaviors was 92.7% (range=85–97%). The third experimenter in Experiment 2 had served as the reliability coder for Experiment 1.

Results

Preliminary Analyses—We performed three sets of preliminary analyses to ensure the validity of the main analyses. First, we used individual *t*-tests to compare performance on the two events used in each test phase. Specifically, we tested for possible differences between events (a) G and H immediately after encoding, (b) A and B at the 15-minute delay, (c) C and D at the 48-hour delay, and (d) E and F at the 2-week delay. There were no statistically significant differences, for either dependent variable: all $ts<.45$, $ps>.50$. For the

main analyses, we used the means of performance on the two event sequences in each test phase.

Second, we tested for differences in encoding of the sequences that would be tested in the different phases. That is, we used ANOVAs to compare immediate recall performance on the means of event sequences AB, CD, EF, and GH. The analyses revealed no difference in immediate recall performance, either in terms of the number of individual target actions produced, $F(3,150)=2.0$, $p=.12$, $\eta_p^2 = .04$, or in terms of the number of pairs of actions produced in target order, $F(3,150)=1.3$, $p=.29$, $\eta_p^2 = .02$. We also compared immediate recall performance on the specific sequences that would be tested after 1 month (i.e., event sequences A, C, E and G). No significant differences emerged, for either the number of individual target actions produced, $F(3,153)=1.4$, $p=.24$, $\eta_p^2 = .03$, or the number of pairs of actions produced in target order, $F(3,153)=1.5$, $p=.23$, $\eta_p^2 = .03$. These analyses confirm that, as expected, there were no differences across sequences at immediate recall.

Third, we tested for differences among events A, C, E, and G at long-term recall, to determine whether the tests administered at 15 minutes, 48 hours, and 2 weeks (events A, C, E, respectively) systematically affected retrieval after the 1-month delay. The analyses revealed no difference in long-term recall, either for the number of individual target actions produced, $F(3,159)=1.5$, $p=.23$, $\eta_p^2 = .03$, or for the number of pairs of actions produced in target order, $Fs(3,159)=1.6$, $p=.19$, $\eta_p^2 = .03$. Thus the intervening tests for post-encoding memory trace status did not systematically affect retrieval performance at long-term recall.

Assessments of Encoding and Retention—Descriptive statistics for infants' performance at immediate recall (encoding), at each of the post-encoding tests (15 minutes, 48 hours, 2 weeks), and at long-term recall are provided in Table 5 (Panel a shows scores across age groups; Panel b shows scores for each age group separately). To determine whether infants retained memories of the event sequences over the delays, we conducted Test (encoding, 15 minutes, 48 hours, 2 weeks, 1 month) × Age (16-, 20-, 24-month-old) mixed measures ANOVAs, with repeated measures on Test, for the number of individual target actions and pairs of actions produced in target temporal order.

Across age groups there was no significant forgetting across delays. For the number of individual target actions produced, $F(4,192)=1.5$, $p=.20$, $\eta_p^2 = .03$, the nonsignificant main effect of Test indicated that performance did not differ reliably at encoding (events GH), 15-minutes after encoding (events AB), 48 hours after encoding (events CD), 2 weeks after encoding (events EF) and 6 weeks after encoding (events ACEG). For the number of pairs of actions produced in target temporal order, the main effect of test approached significance, $F(4,192)=2.4$, $p=.054$, $\eta_p^2 = .05$. As suggested by inspection of Table 5, Panel a, across age groups, there was a gradually decreasing level of performance from encoding to the long-term recall test. Main effects of Age indicated lower levels of performance for 16-month-olds relative to 20- and 24-month-olds, which did not differ from one another, both in terms of the number of individual target actions produced, $F(2,48)=18.2$, $p<.001$, and in terms of production of ordered pairs of actions, $F(2,48)=14.1$ $p<.001$. Neither interaction of Test × Age was significant.

Explanation of Variance in Long-term Recall—To determine the variance in long-term recall accounted for by encoding and by each post-encoding measure of the memory trace, we conducted multiple regression analyses. As in Experiment 1, we entered the measure of encoding trace status (immediate recall) into the model first (Step 1), and then entered the subsequent tests in a step-wise fashion. Within each model, beta weights were evaluated to determine which tests were significant predictors of performance at 1-month.

As can be seen in Table 6, across ages (*N*=51), when it was the sole predictor in the model (Step 1), the measure of encoding accounted for significant variance in both the number of individual target actions and pairs of actions produced in target order. Entry of 15-minute performance into the model as Step 2 increased the variance accounted for from 11% to 38% for individual target actions, and from 11% to 22% for production of ordered pairs of actions. The addition of the 15-minute test also rendered the measure of encoding nonsignificant as a predictor (i.e., only the beta weights for 15-minute performance were significant).

As the third step, we added the measure of performance 48 hours after encoding (Step 3 in Table 6). With the addition, the measure of performance after 15 minutes remained a significant predictor whereas the measure of encoding remained nonsignificant. The addition of 48-hour test performance increased the proportion of variance accounted for from 38% to 55% for individual target actions, and from 22% to 42% for production of ordered pairs of actions. In the final step, we included all four predictors in the models (Step 4). For the number of individual target actions produced, neither immediate nor 2-week recall were significant predictors of long-term recall. In contrast, the measures of memory trace status after 15 minutes and 48 hours remained significant predictors; the total variance accounted for was 56%. For the production of pairs of actions in target order, neither immediate nor 2-week recall were significant predictors of long-term recall. In addition, the contribution of 15-minute recall fell below the level of statistical significance. The measure of memory trace status after 48 hours remained a significant predictor; the total variance accounted for was 43%.

Discussion

The purpose of Experiment 2 was to inform the course of functionally relevant post-encoding changes in trace status. We included tests 15 minutes and 48 hours after encoding, to allow assessment of the memory representation during the initial period of consolidation, and tests 2 weeks after encoding, to provide assessment of the functional significance of the later-stage status of the trace. The 2-week test also allowed us to evaluate an alternate explanation for the results of Experiment 1, namely, that the measure of trace status obtained 1 week after encoding was a stronger predictor of long-term recall, relative to the measure of encoding, because the observations were closer together in time.

Contrary to the alternative hypothesis, the 2-week assessment of trace status was not the strongest predictor of long-term recall, even though it was closer in time to the final assessment. Rather, measures of performance 15 minutes and 48 hours after encoding each accounted for unique variance over and above the variance explained by encoding; the addition of the 2-week assessment did not add significant additional explained variance. For prediction of long-term recall of the individual target actions of the sequences, the 15-minute assessment explained the greatest amount of variance, with additional contribution by the 48-hour assessment. For prediction of long-term recall of the temporal order of the sequences, the 48-hour assessment explained the greatest amount of variance, and in the full model, was the only significant unique predictor. This pattern suggests that item memory (i.e., information about the individual actions of sequences) may stabilize more quickly than order memory (i.e., information about the temporal order of sequences). This suggestion is speculative and requires additional test. At the same time, it is consistent with findings that measures of memory for temporal order are more sensitive to developmental differences in long-term recall, relative to measures of memory for the individual target actions of sequences (e.g., Bauer et al., 2000). Memory for temporal order also is vulnerable to hippocampal compromise (e.g., DeBoer, Wewerka, Bauer, Gerogieff & Nelson, 2005), and is especially reliant on a region of the hippocampus that undergoes a protracted course of development, namely, the dentate gyrus (e.g., Lisman, 1999).

General Discussion

Since the dawn of the 20th century, it has been apparent that memory traces undergo additional processing post encoding. Yet the possibility of post-encoding processes as a source of variance in long-term recall in infancy is just beginning to be explored. The processes have been relatively neglected in spite of evidence of immaturity of the neural structures and network that support post-encoding processes, and thus the distinct possibility that the processes may be a significant source of age-related variance in long-term recall.

The present research extended the small literature on the implications of post-encoding memory processes for long-term recall in infants 16, 20, and 24 months of age. The results of the investigation were clear. When measures of encoding trace status (as measured by immediate recall) were the sole predictors of long-term recall, they explained significant variance. Critically, once measures of post-encoding trace status were entered into the regression models, measures of encoding trace status no longer explained significant variance. Instead, measures of performance 15 minutes (Experiment 2), 48 hours (Experiment 2), and 1 week (Experiment 1) post encoding explained variance in long-term recall. In Experiment 1, total variance in long-term recall explained by trace status 1 week post encoding was 68% and 63% for individual target actions and pairs of actions in temporal order, respectively. In Experiment 2, total variance in long-term recall explained by the combination of 15-minute and 48-hour post-encoding trace status was 56% and 43% for actions and temporal order, respectively. Whether the differences in variance explained in the two experiments are due to the number of predictors (one versus two in Experiments 1 and 2, respectively), to the timing of the post-encoding probes (1 week versus 15 minutes and 48 hours in Experiments 1 and 2, respectively), or to differences in the degree of independence of the observations (measures were wholly and partially dependent in Experiments 1 and 2, respectively), cannot be determined by this research alone. What can be determined is that measures of the status of memory traces minutes to days after encoding predicted significant and unique variance in long-term recall among 16- to 24-month-olds.

One possible alternative explanation for the results of both prior research that has examined post-encoding trace status as a predictor of long-term recall and Experiment 1 was that relative to the measures of encoding trace status, the post-encoding measures were obtained closer to the time of the long-term recall test. More proximate measures would be expected to be better predictors. The results of analyses featuring test of 2-week post-encoding trace status in Experiment 2 speak against this alternative interpretation. That is, though the measure of 2-week post-encoding trace status was obtained closer to the measure of long-term recall, relative to the measures of 15-minute and 48-hour post-encoding trace status, the 2-week delayed recall test was not a unique predictor of long-term recall, whereas the measures obtained closer to the time of encoding predicted substantial variance. In summary, for the tasks used in the present research, and for the age groups represented, tests of memory representations hours to days after encoding were more diagnostic of memory representations 1 month later, compared to test immediately after or weeks after encoding.

The results of the present investigation are consistent with the small body of literature that emphasizes the importance of post-encoding processes for long-term recall in infancy. In some cases, evidence of post-encoding changes in trace status comes in the form of findings of age-related differences in long-term recall in spite of age invariance in encoding (e.g., Bauer, 2005; Howe & Courage, 1997). In other cases, evidence comes in the form of differential changes in memory trace strength over days and weeks as a function of stimulus type, in spite of comparable levels of original encoding (Bauer et al., 2011). Yet other evidence comes from studies such as the present research, in which the absolute and relative contributions of memory trace status at different points in the life of a memory are

differentially predictive of long-term recall (Bauer et al., 2002, 2003). Together, the studies indicate that post-encoding processes play a prominent role in explanation of patterns of remembering and forgetting in infancy.

The present research complements and extends the existing literature that features post-encoding processes as an element in the explanation of variance in long-term recall in infancy. Most importantly in this regard is the information gained regarding the course of functionally relevant post-encoding changes in trace status. Because systematic investigations of the time course of memory trace consolidation have been confined almost exclusively to the animal literature (e.g., Kim & Fanselow, 1992; Takehara et al., 2003; see Bosshardt et al., 2005, for an exception with adult humans), prior to the present research, there was little information upon which to base predictions regarding the most important times at which to interrogate memory traces for their value in predicting long-term recall (see Meeter & Murre, 2004, for discussion of this and related points). The findings of the present research imply that at least under the conditions tested, measurement sooner—rather than later—after encoding, is most predictive. Moreover, there was suggestion in Experiment 2 that item memory may stabilize more rapidly than order memory. What we could not explore in the present research was the possibility of age-related differences in the course of functionally relevant post-encoding changes in trace status. Though the investigation involved three age groups of infants, we did not have sufficient power in the design to examine the patterns of predictive relations separately for each group. Give that the neural structures and network responsible for consolidation processes undergo a protracted course of development (see Seress & Abraham, 2008; Bauer, 2007, 2009, for discussion), it would not be surprising to find that the time course of functionally relevant post-encoding changes in trace status differs across development (see Meeter & Murre, 2004, for similar arguments for differences in the time course of consolidation across species).

We also may expect developmental changes in the amount of age-related variance in long-term recall explained by post-encoding processes. As argued in Bauer (2006), early in development, early-stage processes (encoding and consolidation) account for a large portion of the variance in long-term recall. With development, we may see later-stage processes (retrieval) emerge as the primary determinant of age-related differences in long-term recall. This hypothesis rests on the assumption that with development of the neural substrate implicated in early-stage processes, they will become less susceptible to failure. Though aspects of the hippocampus mature by 6 months (Seress, 2001), the dentate gyrus, frontal cortex, and temporal-cortical connections implicated in consolidation of memory traces develop later. For example, in the dentate gyrus the rise to a maximum level of synapses occurs between 8 and 20 months, and adult levels of synapses are reached at 4–5 years of age (Seress, 2001). For the prefrontal cortex, adult levels of synapses occur in late adolescence, or even early adulthood (Huttenlocher & Dabholkar, 1997). Changes in myelination and in hippocampal and cortical volume also occur up to adolescence (e.g., Giedd et al., 1999; Pfluger et al., 1999). Given that structure and function are related, it is reasonable to expect that as the neural structures that underlie the processes of memory develop, the relative contributions of the processes also will change. Additional research on both behavioral and brain development will be necessary to adequately test this hypothesis.

Though we invoke consolidation processes in the explanation of the patterns of findings in the present research, we acknowledge that we did not measure these processes directly. In healthy human infants, we cannot observe post-encoding processes “at work.” We did not examine the synaptic changes implicated in cellular consolidation or the changes in hippocampal dependence associated with systems consolidation. Further, we did not examine the possibility of reconsolidation processes (i.e., processes by which long-term

memories return to a state of lability as a result of reactivation: e.g., see McKenzie & Eichenbaum, 2011; Bauer, 2009, for discussion). Instead, we used time to make assumptions about the underlying processes. Although we cannot be sure exactly what processes were involved, the results of the present research, coupled with previous studies using neuroimaging (e.g., Bosshardt et al., 2005) and animal models (e.g., Takehara et al., 2003), implicate consolidation processes.

In conclusion, in the present experiments we determined the absolute and relative contributions of encoding and post-encoding processes to explanations of variability in long-term recall in infants in the second year of life. Measures of encoding trace status and measures of post-encoding trace status each contributed significant variance when considered alone. Importantly, in combination, only measures of post-encoding trace status predicted significant unique variance in long-term recall. The amount of variance explained was maximized when assessments of trace status were 15 minutes to 1 week post encoding. The findings have implications for explanations of the causes of age-related and individual variability in long-term recall in infancy. They also compel future research involving multiple age groups of infants and children, and finer grained assessments of the time course of functionally relevant changes in post-encoding trace status. Such research will help to further chart the development of memory, as well as further elaborate our explanations of it.

Acknowledgments

Support for this research was provided by NIH grants (HD28425 and HD42483) to Patricia J. Bauer, and by Emory College of Arts and Sciences. The authors thank Lindsay Lewis, Kevin Dugas, Alisha Holland, and Kristen Parker for help with data collection, and members of the Memory at Emory laboratory group for help at various stages of the research. The authors extend a special note of gratitude to the infants and families who participated in the research.

References

Adlam AL, Vargha-Khadem F, Mishkin M, de Haan M. Deferred imitation of action sequences in developmental amnesia. Journal of Cognitive Neuroscience. 2005; 17:240–248. [PubMed: 15811236]

Barr R, Dowden A, Hayne H. Developmental changes in deferred imitation by 6–24-monthold infants. Infant Behavior and Development. 1996; 19:159–170.

Bauer PJ. Holding it all together: How enabling relations facilitate young children's event recall. Cognitive Development. 1992; 7:381–401.

Bauer PJ. Developments in declarative memory: Decreasing susceptibility to storage failure over the second year of life. Psychological Science. 2005; 16:41–47. [PubMed: 15660850]

Bauer PJ. Getting explicit memory off the ground: Steps toward construction of a neuro-developmental account of changes in the first two years of life. Developmental Review. 2004; 24:347–373.

Bauer PJ. Constructing a past in infancy: A neuro-developmental account. Trends in Cognitive Sciences. 2006; 10:175–181. [PubMed: 16537115]

Bauer, PJ. Remembering the times of our lives: Memory in infancy and beyond. Mahwah, NJ: Erlbaum; 2007.

Bauer PJ. Toward a neuro-developmental account of the development of declarative memory. Developmental Psychobiology. 2008; 50:19–31. [PubMed: 18085555]

Bauer, PJ. Neurodevelopmental changes in infancy and beyond: Implications for learning and memory. In: Barbarin, O.; Wasik, B., editors. Handbook of developmental science and early education. New York: Guilford Publications, Inc; 2009. p. 78-102.

Bauer, PJ. Memory. In: Zelazo, PD., editor. Oxford handbook of developmental psychology. New York: Oxford University Press; (in press).

Bauer, PJ.; Cheatham, CL.; Cary, MS.; Van Abbema, DL. Short-term forgetting: Charting its course and its implications for long-term remembering. In: Shohov, SP., editor. Advances in psychology research. Vol. Volume 9. Huntington, NY: Nova Science Publishers; 2002. p. 53-74.

Bauer PJ, Dow GAA. Episodic memory in 16- and 20-month-old children: Specifics are generalized, but not forgotten. Developmental Psychology. 1994; 30:403–417.

Bauer PJ, Hertsgaard LA. Increasing steps in recall of events: factors facilitating immediate and long-term memory in 13.5- and 16.5-month-old children. Child Development. 1993; 64:1204–1223. [PubMed: 8404265]

Bauer PJ, Güler OE, Starr RM, Pathman T. Equal learning does not result in equal remembering: The importance of post-encoding processes. Infancy. 2011; 16:557–586.

Bauer PJ, Mandler JM. One thing follows another: Effects of temporal structure on 1- to 2-year-olds' recall of events. Developmental Psychology. 1989; 25:197–206.

Bauer PJ, Travis LL. The fabric of an event: Different sources of temporal invariance differentially affect 24-month-olds' recall. Cognitive Development. 1993; 8:319–341.

Bauer PJ, Wenner JA, Dropik PL, Wewerka SS. Parameters of remembering and forgetting in the transition from infancy to early childhood. Monographs of the Society for Research in Child Development. 2000; 65(Serial No. 363)

Bauer PJ, Wiebe SA, Carver LJ, Waters JM, Nelson CA. Developments in long-term explicit memory late in the first year of life: Behavioral and electrophysiological indices. Psychological Science. 2003; 14:629–635. [PubMed: 14629697]

Bosshardt S, Degonda N, Schmidt CF, Boesiger P, Nitsch RM, Hock C, Henke K. One month of human memory consolidation enhances retrieval-related hippocampal activity. Hippocampus. 2005; 15:1026–1040. [PubMed: 16015623]

Burch MM, Schwade JA, Bauer PJ. Finding the right fit: examining developmentally appropriate levels of challenge in elicited imitation studies. Advances in Child Development and Behavior. 2010; 38:27–48. [PubMed: 21207804]

DeBoer T, Wewerka S, Bauer PJ, Georgieff MK, Nelson CA. Explicit memory performance in infants of diabetic mothers at 1 year of age. Developmental Medicine and Child Neurology. 2005; 47:525–531. [PubMed: 16108452]

Dudai, Y.; Morris, RGM. To consolidate or not to consolidate: what are the questions?. In: Bolhuis, JJ., editor. Brain, Perception, Memory. Advances in Cognitive Sciences. Oxford: Oxford University Press; 2000.

Eichenbaum, HE.; Cohen, NJ. From conditioning to conscious recollection: Memory systems of the brain. New York, NY: Oxford University Press; 2001.

Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL. Brain development during childhood and adolescence: A longitudinal MRI study. Nature Neuroscience. 1999; 2:861–863.

Hayne H. Infant memory development: Implications for childhood amnesia. Developmental Review. 2004; 24:33–73.

Howe ML, Brainerd CJ. Development of children's long-term retention. Developmental Review. 1989; 9:301–340.

Howe ML, Courage ML. On resolving the enigma of infantile amnesia. Psychological Bulletin. 1993; 113:305–326. [PubMed: 8451337]

Howe ML, Courage ML. Independent paths in the development of infant learning and forgetting. Journal of Experimental Child Psychology. 1997; 67:131–163. [PubMed: 9388803]

Huttenlocher PR, Dabholkar AS. Regional differences in synaptogenesis in human cerebral cortex. The Journal of Comparative Neurology. 1997; 387:167–178. [PubMed: 9336221]

Lisman J. Relating Hippocampal Circuitry to Function: Recall of Memory Sequences by Reciprocal Dentate/CA3 Interactions. Neuron. 1999; 22:233–242. [PubMed: 10069330]

Kim JJ, Fanselow MS. Modality-specific retrograde amnesia of fear. Science. 1992; 256:675–677. [PubMed: 1585183]

McDonough L, Mandler JM, McKee R, Squire L. The Deferred Imitation Task as a Nonverbal Measure of Declarative Memory. Proceedings of the National Academy of Sciences. 1995; 92:7580–7584.

McGaugh JL. Memory - A century of consolidation. Science. 2000; 287:248–251. [PubMed: 10634773]

Meeter M, Murre JMJ. Consolidation of long-term memory: Evidence and alternatives. Psychological Bulletin. 2004; 130:843–857. [PubMed: 15535740]

McKenzie S, Eichenbaum H. Consolidation and reconsolidation: two lives of memories? Neuron. 2011; 71:224–233. [PubMed: 21791282]

Müller GE, Pilzecker A. Experimentelle Beitrage zur Lehre vom Gedächtnis. Zeitschrift für Psychologie. Ergänzungsband. 1900; 1:1–300.

Nelson, CA.; de Haan, M.; Thomas, K. D. Kuhn & R. Siegler (Volume Editors: Volume 2—Cognition, Perception, and Language), W. Damon & R. M. Lerner (Editors-in-Chief).Handbook of Child Psychology, Sixth Edition. Hoboken, NJ: John Wiley & Sons, Inc.; 2006. Neural bases of cognitive development; p. 3-57.

Pfluger T, Weil S, Wies S, Vollmar C, Heiss D, Egger J, Scheck R, Hahn K. Normative volumetric data of the developing hippocampus in children based on magnetic resonance imaging. Epilepsia. 1999; 40:414–423. [PubMed: 10219266]

Rose SA, Feldman JF, Jankowski JJ. Infant visual recognition memory. Developmental Review. 2004; 24:74–100.

Rose SA, Gottfried AW, Melloy-Carminar P, Bridger WH. Familiarity and novelty preferences in infant recognition memory: Implications for information processing. Developmental Psychology. 1982; 18:704–713.

Seress, L. Morphological changes of the human hippocampal formation from midgestation to early childhood. In: Nelson, AA.; Luciana, M., editors. Handbook of developmental cognitive neuroscience. Cambridge, MA: MIT Press; 2001. p. 45-47.

Seress, L.; Abraham, H. Pre- and postnatal morphological development of the human hippocampal formation. In: Nelson, CA.; Luciana, M., editors. Handbook of developmental cognitive neuroscience, second edition. Cambridge, MA: The MIT Press; 2008. p. 187-212.

Squire LR, Alvarez P. Retrograde amnesia and memory consolidation: a neurobiological perspective. Current Opinion in Neurobiology. 1995; 5:169–177. [PubMed: 7620304]

Takehara K, Kawahara S, Kirino Y. Time-dependent reorganization of the brain components underlying memory retention in trace eyeblink conditioning. Journal of Neuroscience. 2003; 23:9897–9905. [PubMed: 14586019]

Wiebe SA, Bauer PJ. Interference from additional props in an elicited imitation task: When in sight, firmly in mind. Journal of Cognition and Development. 2005; 6:325–364.

Wixted JT. The psychology and neuroscience of forgetting. Annual Review of Psychology. 2004; 55:235–269.

Zola, SM.; Squire, LR. The medial temporal lobe and the hippocampus. In: Tulving; Craik, editors. The Oxford Handbook of Memory. Oxford University Press; 2000. p. 501-520.

Highlights

- We tested the contribution of encoding and post-encoding processes to recall in infancy
- The status of 16–24-mo-olds' memory representations were tested at several time points
- Alone, measures of encoding and post-encoding trace status predicted long-term recall
- In combination, only measures of post-encoding trace status predicted long-term recall
- Implications for explanations of causes of variability in long-term recall discussed

Table 1

Experiment 1: The Number of Participants who Reached Criterion on 0, 1, 2, or All 3 Event Sequences in the Criterion Encoding Condition

Number of Sequences on which Participant Reached Criterion at Encoding	Age Group		
	16-month-olds	20-month-olds	24-month-olds
0 (infant excluded from analyses)	6	2	3
1 (1 sequence included in analyses)	6	3	10
2 (*M* of 2 sequences included in analyses)	4	10	2
3 (*M* of 3 sequences included in analyses)	1	2	2

Table 2

Experiment 1: Mean Number (and Standard Deviation) of Actions and Pairs of Actions in Target Temporal Order for Each Age Group and Test for the Criterion (Panel a) and Standard (Panel b) Encoding Conditions

Age/ Dependent Measure	Test			
	Baseline	Immediate	1 Week	1 Month
	M (SD)	M (SD)	M (SD)	M (SD)
Panel a: Criterion Encoding Condition				
16-month-olds				
Actions (max=2)	.61 (.47)	NA	1.70 (.51)	1.55 (.65)
Pairs (max=1)	.03 (.10)	NA	.62 (.46)	.55 (.47)
20-month-olds				
Actions (max=3)	.98 (.39)	NA	2.70 (.47)	2.53 (.62)
Pairs (max=2)	.11 (.21)	NA	1.54 (.49)	1.32 (.66)
24-month-olds				
Actions (max=4)	2.06 (.48)	NA	3.80 (.37)	3.81 (.31)
Pairs (max=3)	.55 (.41)	NA	2.67 (.45)	2.49 (.67)
Panel b: Standard Encoding Condition				
16-month-olds				
Actions (max=2)	.80 (.49)	1.37 (.48)	1.35 (.46)	1.37 (.42)
Pairs (max=1)	.14 (.21)	.53 (.31)	.47 (.31)	.45 (.31)
20-month-olds				
Actions (max=3)	1.29 (.44)	2.55 (.41)	2.08 (.70)	1.92 (.69)
Pairs (max=2)	.29 (.35)	1.41 (.36)	.98 (.61)	.90 (.62)
24-month-olds				
Actions (max=4)	1.45 (.68)	3.39 (.54)	3.27 (.60)	2.96 (.90)
Pairs (max=3)	.41 (.45)	2.20 (.46)	1.96 (.47)	1.65 (.70)

Table 3

Experiment 1: Immediate and 1-week Recall as Predictors of 1-month Recall in the Criterion (Panel a) and Standard (Panel b) Encoding Conditions

			Individual Actions			Actions in Target Temporal Order		
Dependent Variable (Predicted)			**B**	**SE B**	**β**	**B**	**SE B**	**β**
Panel a: Criterion Encoding Condition								
1-month Recall			.97	.09	.87 ***	.83	.11	.79 ***
			$F(1,38)=123.24$, $p<.0001$, adj. $R^2=.76$			$F(1,38)=62.93$, $p<.0001$, adj. $R^2=.61$		
Panel b: Standard Encoding Condition								
1-month Recall	Step 1	Immediate	.73	.10	.73 ***	.70	.09	.73 ***
			$F(1,49)=56.05$, $p<.0001$, adj. $R^2=.52$			$F(1,49)=56.49$, $p<.0001$, adj. $R^2=.53$		
	Step 2	Immediate	.04	.16	.04	.18	.16	.18
		1-week	.77	.15	.80 ***	.62	.16	.64 ***
			$F(2,48)=54.79$, $p<.0001$, adj. $R^2=.68$, $\Delta R^2=.16$			$F(2,48)=43.97$, $p<.0001$, adj. $R^2=.63$, $\Delta R^2=.11$		

* $p<.01$;

** $p<.001$;

*** $p<.0001$

Table 4

Experiment 2: Schematic Design

Modeling of Events (Encoding)	Test Immediately after Encoding	Post-encoding Tests			Long-term Memory Retrieval
	Time 0	15 minutes	48 hours	2 weeks	6 weeks
A	A	A			A
B*	B	B			
C	C		C		C
D	D		D		
E	E			E	E
F*	F			F	
G	G				G
H*	H				

Note. Events marked with an asterisk were introduced at Session 2. All other events were introduced at Session 1.

J Exp Child Psychol. Author manuscript; available in PMC 2014 February 01.

Table 5

Experiment 2: Mean Number (and Standard Deviation) of Actions and Pairs of Actions in Target Temporal Order across Age Groups (Panel a) and for Each Age Group Separately (Panel b) for Each Test

Age/ Dependent Measure	Test				
	Immediate	**15 minutes post encoding**	**48 hours post encoding**	**2 weeks post encoding**	**6 weeks post encoding**
	M (SD)	**M (SD)**	**M (SD)**	**M (SD)**	**M (SD)**
Panel a: Across age groups					
Actions	2.26 (.79)	2.07 (.93)	2.09 (.77)	2.05 (.79)	1.98 (.62)
Pairs	1.20 (.69)	1.06 (.68)	1.05 (.63)	.95 (.69)	.91 (.52)
Panel b: Age groups separately					
16-month-olds					
Actions	1.79 (.83)	1.44 (.75)	1.79 (.77)	1.68 (.58)	1.51 (.68)
Pairs	.79 (.69)	.65 (.52)	.79 (0.64)	.62 (.49)	.57 (.56)
20-month-olds					
Actions	2.38 (0.78)	1.91 (0.97)	2.06 (0.79)	2.15 (0.96)	2.04 (0.47)
Pairs	1.29 (.73)	.97 (.82)	1.06 (.55)	1.03 (.84)	.99 (.44)
24-month-olds					
Actions	2.62 (.52)	2.88 (.22)	2.41 (.64)	2.32 (.66)	2.37 (.35)
Pairs	1.50 (.43)	1.56 (.24)	1.29 (.61)	1.21 (.59)	1.16 (.38)

Note. Maximum for individual target actions=3.00 and for pairs of actions in target temporal order=2.00.

Table 6

Experiment 2: Immediate, 15-minute, 48-hour, and 2-week Recall as Predictors of Long-term-month Recall

		Individual Actions			Actions in Target Temporal Order		
Entered into Model	**Predictor**	**B**	**SE B**	**β**	**B**	**SE B**	**β**
Step 1	Immediate	.28	.11	.36*	.27	.10	.35*
		$F(1,50)=7.07$, p<.02, adj. R^2=.11			$F(1,50)=6.94$, p<.02, adj. R^2=.11		
Step 2	Immediate	.04	.10	.05	.10	.11	.13
	15-minute test	.41	.09	.61***	.32	.11	.42**
		$F(2,50)=16.43$, p<.0009, adj. R^2=.38, ΔR^2=.28			$F(2,50)=7.88$, p<.002, adj. R^2=.22, ΔR^2=.12		
Step 3	Immediate	.03	.09	.03	.07	.10	.09
	15-minute test	.33	.08	.50***	.25	.10	.32*
	48-hour test	.35	.08	.43***	.39	.09	.47***
		$F(3, 50)=21.28$, p<.0009, adj. R^2=.55, ΔR^2=.17			$F(3,50)=12.84$, p<.0009, adj. R^2=.42, ΔR^2=.20		
Step 4	Immediate	.05	.09	.06	.08	.10	.10
	15-minute test	.29	.08	.44**	.19	.11	.25+
	48-hour test	.31	.08	.39***	.36	.10	.43***
	2-week test	.12	.09	.15	.14	.09	.18
		$F(4, 50)=16.80$, p<.0009, adj. R^2=.56, ΔR^2=.02			$F(4,50)=10.43$, p<.0009, adj. R^2=.43, ΔR^2=.03		

+ p<.10,

* p<.05;

** p<.009;

*** p<.0009