Published in final edited form as:

Infancy. 2011 ; 16(6): 557–586. doi:10.1111/j.1532-7078.2010.00057.x.

Equal Learning Does Not Result in Equal Remembering: The Importance of Post-Encoding Processes

Patricia J. Bauer,

Department of Psychology, Emory University

O. Evren Güler,

Center for Neurobehavioral Development, University of Minnesota

Rebecca M. Starr, and

Child Trends

Thanujeni Pathman

Department of Psychology, Emory University

Abstract

Explanations of variability in long-term recall typically appeal to encoding and/or retrieval processes. However, for well over a century, it has been apparent that for memory traces to be stored successfully, they must undergo a post-encoding process of stabilization and integration. Variability in post-encoding processes is thus a potential source of age-related and individual variance in long-term recall. We examined post-encoding variability in each of two experiments. In each experiment, 20-month-old infants were exposed to novel three-step sequences in each of three encoding conditions: watch only, imitate, and learn to criterion. They were tested for recall after 15 min (as a measure of the success of encoding) and either weeks (1, 2, or 3: Experiment 1) or days (1, 2, or 4: Experiment 2) later. In each experiment, differential relative levels of performance among the conditions were observed at the two tests. The results implicate post-encoding processes as a source of variance in long-term recall.

Since the dawn of the 20th century (Müller & Pilzecker, 1900), it has been apparent that what is remembered about an event or episode is determined not only by how successfully the event was encoded to begin with and how successfully the memory trace was retrieved, but also by how successfully the encoded material was consolidated. Consolidation is the process by which initially labile memory traces are stabilized and integrated into long-term storage (see McGaugh, 2000; Wixted, 2004, for reviews). The process is subserved by a particular neural substrate (e.g., Eichenbaum & Cohen, 2001), that undergoes substantial postnatal developmental change throughout infancy and into the school years (e.g., Bauer, 2006; Nelson, 1997). This makes it an attractive candidate as a source of age-related and individual variability in long-term recall (Bauer, 2009a, 2009b). Yet to date, few studies have examined the integrity of memory traces during the period in which they undergo consolidation, namely, days and weeks after experience of an event. As a result, it is difficult

to evaluate the possibility that differential post-encoding contributes to variance in long-term recall. The purpose of this research was to take some initial steps in this chain of inquiry.

The process of consolidation originally was hypothesized by Müller and Pilzecker (1900) to account for retroactive interference. In laboratory tests they observed that new material learned shortly after—but not long after—old material produced deficits in memory for the old material. For example, they examined memory for lists of words in two situations. In one, List 1 was learned, time was allowed to pass, and then List 2 was learned, followed by tests of memory for both lists. In this situation, high levels of memory for both lists were observed. In the other situation, List 1 was learned, and then very shortly thereafter List 2 was presented. In this case, they observed good recall of List 2, but poor recall of List 1. It seemed that in the short-delay situation, List 2 retroactively interfered with List 1. Müller and Pilzecker advanced the hypothesis that there was retroactive interference because at the time List 2 was learned, List 1 had not yet been stabilized or integrated into storage, a process they termed *consolidation*.

Müller and Pilzecker's (1900) work served to illustrate an important principle about memory, namely, that processes that take place after encoding influence later remembering. This principle is further supported by patient work as well as work with animal models of lesion and disease. One source of evidence for the importance of post-encoding processes is from individuals with medial temporal lesions who consequently suffer anterograde amnesia: an inability to form new declarative (or explicit) memories. Patients with damage in the medial temporal lobe have normal intelligence (as measured by standardized IQ tests), normal short-term memory (e.g., over intervals of a few seconds, they can remember a series of digits as well as healthy control subjects), and preserved memory for events from the distant past. However, they perform at levels below normal control participants on a variety of memory tasks that require new learning, including reproducing a diagram after a 5- to 10-min delay, recalling and recognizing individual words presented on lists, and recognizing words and faces (e.g., Reed & Squire, 1998). The memory deficits of these individuals cannot be accounted for by problems with encoding alone: they have preserved short-term or immediate memory. Nor can they be explained by problems with retrieval alone: lower levels of performance are apparent on tests of recognition as well as recall, even though tests of recognition make lower retrieval demands, relative to the demands of recall. These observations imply that for new memories to be effectively stored, they must undergo additional processing after initial registration.

The second source of evidence that to live on, memories must undergo post-encoding processing is the observation of temporally graded retrograde amnesia: Memory for more recent events is impaired, relative to memory for more remote events (see Brown, 2002, for a review). Notice that this pattern is precisely the opposite of normal forgetting. The phenomenon is observed in humans whose amnesia was the result of Korsakoff's syndrome (attributed to chronic alcohol abuse) as well as patients whose amnesia was acute, because of lesion, infarction, or anoxia. It also can be induced in nonhuman animals (including rabbits, mice, and monkeys: see Eichenbaum & Cohen, 2001; Squire & Alvarez, 1995, for reviews) by creating a lesion in medial temporal structures at different points after learning of a novel association (e.g., association between a tone and an electrical shock). Lesions

made shortly after learning produce a large deficit in performance; lesions made well after training produce only mild or no disruption of performance (e.g., Kim & Fanselow, 1992; Takehara, Kawahara, & Kirino, 2003). Together, the data on temporally graded retrograde amnesia and on anterograde amnesia provide strong evidence that for memories to be preserved over the long term, they must undergo additional processing for some time after experience of an event.

As argued by Bauer (2006, 2009a, 2009b), there is reason to believe that the post-encoding processes that transform labile patterns of neural activation into stable memory traces may be especially challenging for infants and very young children, because of immaturity of the neural structures and thus the cognitive processes implicated in the transformation.¹ Briefly, consolidation of new memory traces is considered to involve two processes that occur in parallel: (a) stabilization of the trace through formation of associations among the individual elements of experience, and (b) integration of the trace into cortical association areas (e.g., Zola & Squire, 2000). As summarized in Bauer (2007, 2009b), it is understood that stabilization of a memory trace begins as the inputs from distributed cortical association areas are projected to structures in the medial temporal lobes. The neural codes of the representations of these inputs come together in the perirhinal and parahippocampal cortices of the medial temporal lobes. These cortices then relay the information to the entorhinal cortex, which in turn relays it into the hippocampus itself (by way of the dentate gyrus). It is in the hippocampus that enduring links between the different elements of experience are forged. Even as it is being processed in the hippocampus and surrounding temporal cortices, new information is being associated with old information in cortical storage areas through synchronous convergence, or simultaneous activation of shared elements: neurons that are repeatedly activated together tend to become associated. The result is an entire pattern of interconnection of new information with old (e.g., Eichenbaum & Cohen, 2001; Kandel & Squire, 2000; McGaugh, 2000; Zola & Squire, 2000).

In the developing human, these processes may be expected to be less efficient and effective, because of the relative immaturity of the neural structures and network on which they depend. Portions of the medial temporal structures, including the cell fields of the hippocampus, mature relatively early (e.g., Seress & Abraham, 2008). In contrast, prefrontal cortex and the dentate gyrus of the hippocampus are later to mature. It is not until 20–24 months that the numbers of synapses in these structures peak (Huttenlocher & Dabholkar, 1997), heralding their functional maturity (Goldman-Rakic, 1987). It is not until late in the preschool years and adolescence or early adulthood that adult numbers of synapses are apparent in the dentate gyrus and prefrontal cortex, respectively (Bourgeois, 2001; Huttenlocher & Dabholkar, 1997), indicating full maturity of the structures (Goldman-Rakic, 1987). The connections between the structures also are slow to develop (e.g., Schneider, Il'yasov, Hennig, & Martin, 2004). Late development of cortical structures is important because they are implicated in all phases of the life of a memory. Late

---

¹This argument applies to consolidation processes at the level of the neural structures and systems involved. It may not extend to short-term consolidation-associated events at the cellular and molecular levels, including changes in gene expression, protein synthesis, neurotransmitter release, and synapse formation. These processes are highly conserved across species (e.g., Eichenbaum & Cohen, 2001; Kandel & Squire, 2000; McGaugh, 2000), and likely also across ontogeny.

development of the dentate gyrus is critical because in the mature organism, it is the major “route in” to the hippocampus. Less efficient and effective communication between cortical structures and the hippocampus would present challenges to post-encoding processes and thus storage of new information (Gluck & Myers, 2001; see Bauer, 2006, 2007, 2009a, 2009b, for reviews).

In the literature on the development of memory in the human infant, it has been difficult to evaluate the efficacy of post-encoding processes because of features of the research designs that have been employed to test recall. Major requirements of an adequate test of post-encoding processes are a measure of encoding and a measure of the strength of the memory trace thereafter. Moreover, it is important that encoding be controlled or otherwise equalized across groups or conditions. When encoding varies, it is impossible to determine whether differential levels of recall are because of the differential post-encoding processes or preexisting differences in encoding. These requirements are not often met in the infancy literature, due to the prevailing use of the technique of deferred imitation (Meltzoff, 1985). In deferred imitation, props are used to produce an action or a sequence of actions that the infant observes (and presumably encodes) and then is encouraged to reproduce at a later time (e.g., Barr, Dowden, & Hayne, 1996; Hayne, Boniface, & Barr, 2000; Klein & Meltzoff, 1999; Liston & Kagan, 2002; Meltzoff, 1988a). Because the measure of memory is obtained only after a delay, there is no means to assess how well the material was encoded. As a result, studies that rely exclusively on deferred imitation are not useful for evaluation of the efficacy of post-encoding processes.

The few studies that permit assessment of the possibility of differential post-encoding processes are consistent with the suggestion that they are a significant source of variance in long-term recall. One example is Bauer (2005) in which infants of different ages were matched for levels of encoding prior to imposition of 1- to 6-month delays. In Experiment 1, 13- and 16-month-olds were matched for encoding of three-step sequences; in Experiment 2, 16- and 20-month-olds were matched for encoding of four-step sequences. Although the groups were matched prior to imposition of the delays, there were age-related differences in long-term recall; age-related differences in forgetting were especially apparent at the longer delays. The possibility that the age-related differences were because of differential retrieval success was assessed by measuring relearning. That is, after the test for long-term recall, the sequences were remodeled and the infants were given the opportunity to imitate. Even with the retrieval burden substantially lessened in this way, in both experiments, older infants showed higher levels of performance. A second example is Howe and Courage (1997) in which 15- and 18-month-old (Experiment 1) and 12- and 15-month-old (Experiment 2) infants were brought to criterion levels of learning (and thus, encoding) prior to imposition of a 3-month delay. After the delay, the 12-month-olds remembered less than the 15-month-olds; the 15-month-olds remembered less than the 18-month-olds. Relative levels of performance did not change over the course of four test trials, suggesting storage failure rather than retrieval failure on the part of the younger infants. In both of these studies (Bauer, 2005; Howe & Courage, 1997), in spite of equal encoding, there were age-related differences in long-term recall.

This research is a complement to and extension of the small literature that permits assessment of post-encoding processes. We elected to assess the possibility of differential post-encoding processes within, rather than between-subjects. A within-subjects test is especially powerful because it reduces the plausibility of differentially effective retrieval as an explanation for differential levels of performance after a delay. Even when steps are taken to eliminate retrieval failure as an explanation for between-group differences in performance (as in Bauer, 2005; Howe & Courage, 1997), the possibility remains. In contrast, assuming equal encoding and comparable testing conditions, in a within-subjects comparison, it is more challenging to explain differential recall in terms of differences in retrieval success, except as stemming from differential post-encoding processes and storage failure.

We accomplished the within-subjects test by arranging three different learning experiences that have been associated with different levels of long-term recall. Specifically, prior research has shown that infants who only watch sequences modeled (for whom imitation is deferred) have lower levels of long-term recall performance relative to infants who are permitted to imitate prior to imposition of delays (e.g., Hayne, Barr, & Herbert, 2003; Lukowski et al. 2005; although see Barr & Hayne, 1996; Bauer, Hertsgaard, & Wewerka, 1995). These studies cannot be brought to bear on the questions motivating this research because there were no measures of encoding in the deferred condition and thus no means of determining whether differential performance after a delay was because of differential loss of information over time, or differential encoding. Accordingly, in this research we included both “imitate” and “watch” conditions. In the imitate condition, the measure of encoding was performance immediately after modeling. In the watch condition, the measure of encoding was performance 15 min after modeling. Prior research with typically developing infants has shown no decrement in performance 10–15 min after modeling relative to immediate imitation conditions (i.e., differences between imitate and watch conditions emerge with delays between encoding and test: e.g., Bauer, Van Abbema, & de Haan, 1999; Cheatham, Sesma, Bauer, & Georgieff, in press). Thus, the brief delay permitted instantiation of the encoding condition manipulation while simultaneously affording a valid proxy for the success of encoding. Finally, we included a “criterion” condition in which infants were required to demonstrate errorless reproduction of the sequences, thus ensuring complete encoding (Howe & Courage, 1997). We reasoned that this condition should support especially robust long-term recall.

Between the two experiments we examined recall weeks after encoding (Experiment 1) and days after encoding (Experiment 2). Because post-encoding processes have received little research attention in the adult (Wixted, 2004) and developmental (Bauer, 2009b) literatures, we had little to inform selection of the timepoint (or points) at which differential processing might become apparent. To begin to inform this question, in Experiment 1, we sampled memory at 1, 2, and 3 weeks after encoding. Work with animal models indicates that by 14–28 days post-encoding, memories 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). To preview the results of Experiment 1, we found differential patterns of performance minutes after learning and weeks later, but no changes in relative levels of performance between 1 and 3 weeks after learning. To more precisely

pinpoint the timing of changes in relative levels of performance, in Experiment 2, we sampled memory 1, 2, and 4 days after encoding. Previous research led us to expect differences in performance after these delays, relative to immediately after learning (e.g., Bauer et al., 1999; Hanna & Meltzoff, 1993; Meltzoff, 1988b; Tomasello, Savage-Rumbaugh, & Kruger, 1993).

In both experiments, the participants were 20 months of age. We selected this age because by this point in development, infants competently perform the imitation task under no-delay conditions (see Bauer, 1997, for a review). This suggests that they are successfully encoding the events, a prerequisite to tests of memory for them. All of the event sequences were three steps in length. Twenty-month-old infants perform competently on three-step event sequences (see Bauer, 1997; Burch, Schwade, & Bauer, in press, for reviews). Following the logic just outlined, this therefore provided reasonable assurance that the events would be adequately encoded. Finally, all of the test sequences were constrained by enabling relations: In order to reach a particular outcome or goal, one action in the sequence was both prior to and necessary for a subsequent action (see Bauer, 1992, for discussion). This choice was made because at 20 months of age, infants do not yet perform reliably on test sequences that lack this temporal structure (Bauer, Hertsgaard, Dropik, & Daly, 1998; Bauer & Thal, 1990; Wenner & Bauer, 1999).

Experiment 1: Minutes and Weeks

Method

Participants—The participants were 60 infants with a mean age of 20 months, 3 days (range = 19 months 18 days to 20 months 8 days). Half of the infants were girls. The infants were drawn from a pool of potential participants whose parents had indicated interest in research at the time of their infants' births. The sample was predominantly Caucasian and drawn from families of middle socioeconomic status. All infants participated in two laboratory sessions separated by 1, 2, or 3 weeks (M delays between test sessions = 7, 14.07, and 21.15 days, respectively; ranges = 7–7, 14–15, and 21–23 days, respectively). Infants were pseudo-randomly assigned to delay conditions, with the provision that an equal number of infants was assigned to each delay condition ($N = 20$), and each delay condition included an equal number of girls and boys. At the end of the second session all infants received a small toy for their participation.

Materials—The stimuli were 12 sequences, each of which was three steps in length. All sequences were constrained by enabling relations, such that one action in a sequence is temporally prior to and necessary for the next action. All sequences were novel to the children (based on parent report) and had been used in prior studies (e.g., Bauer & Mandler, 1989; Bauer et al., 1999; Bauer, Wenner, Dropik, & Wewerka, 2000; Wiebe & Bauer, 2005). An example sequence is “make a gong” by folding a bar across a swing-set shaped base to form a support; hanging a metal disk from the bar; and hitting the metal disk with a small mallet, thus causing it to “gong.” A complete list of the sequences used is provided in the Appendix.

Procedure—The sessions took place in a laboratory playroom and were video-taped. All sessions were conducted by one of two female experimenters. Infants were tested by the same experimenter at both sessions. The experimenters tested an approximately equal number of infants in each delay condition. Prior to testing, the experimenters practiced to ensure uniform administration of the protocol. Fidelity to experimental procedures was checked regularly throughout data collection through reviews of the videotapes of the sessions.

After parents gave informed consent for their infants' participation, infants were seated across an adult-sized table from an experimenter, either in their parents' laps or in a booster seat. The experimenter engaged the infant in warm-up play with commercially available toys (beads and a bucket and a slinky and ball) until the experimenter determined the infant to be comfortable with the setting.

All infants participated in three encoding conditions (i.e., encoding condition was within-subjects): watch, imitate, and learn to criterion. They were exposed to and tested on four sequences in each condition. Two sequences in each condition were used to assess the success of encoding and two sequences in each condition were used to assess the status of the trace during the period of consolidation (see below).

The procedure for the watch condition was typical of deferred imitation paradigms. That is, the sequences were modeled twice in succession with simple narration (e.g., “make a gong”: “Watch how I make a gong. Put on the bar. Hang up the bell. Ring it.”). Infants were not given the opportunity for immediate reproduction of the sequences (see Table 1, Panel a). The procedure for the imitate condition was typical of elicited imitation paradigms. That is, the sequences were modeled twice in succession with simple narration, after which infants were given the opportunity for immediate reproduction of the sequences. In the criterion condition, as in Howe and Courage (1997), infants were required to reproduce the sequences in the correct temporal order two times in succession. As in the watch and imitate conditions, the sequences initially were modeled twice in succession. As in the imitate condition, immediately after the second modeling, infants were given the opportunity to reproduce the sequence. The sequences then were modeled again, and infants were given a second opportunity to imitate. If infants reproduced the sequence in correct temporal order on both imitation trials, the procedure was terminated. If not, modeling followed by imitation was repeated a maximum of two more times, until infants achieved two errorless reproductions in a row, or until the sequence had been modeled a total of five times (two initial modelings followed by three additional learning trials). The procedure was terminated after five modelings in order to avoid frustration or excessive fatigue on the part of the infant. For all encoding conditions, the lengths of the response periods were “child controlled,” as defined in prior research using elicited imitation (e.g., Bauer et al., 2000; Howe & Courage, 1997): a response period ended when the infant had produced all of the target actions, pushed the props back to the experimenter, or began engaging in repetitive or exploratory behavior. Sequences were counterbalanced such that each was used approximately equally often in each condition (watch, imitate, criterion).

As a measure of how well the sequences had been encoded, half of the sequences in each condition (two sequences per condition) were tested 15 min after presentation. To measure how well the sequences were being consolidated, the other half of the sequences in each condition were tested either 1, 2, or 3 weeks later (the length of the weeks delay was between-subjects). For the delayed tests, the materials for each sequence in turn were given to the infant with the general prompt: “What can you do with that stuff?” This prompt was followed by a specific reminder in the form of the name of the sequence: “With this stuff you can make a gong. Show me how you make a gong.”

As depicted in Table 1, Panel b, infants were exposed to nine of the 12 sequences at their first visit to the laboratory (designated as Sequences A–I in Table 1, Panel b): three sequences used to measure the success of encoding (one sequence in each encoding condition; Sequences A, D, G), and the six sequences used to measure the success of consolidation (two sequences in each encoding condition; Sequences B, C, E, F, H, I). The remainder of the sequences used to measure the success of encoding (one sequence in each encoding condition; Sequences J, K, L) were presented and tested at the second visit to the laboratory. The decision to test half of the encoding-test sequences at each session was made in order not to overwhelm the infants at the first visit (all six of the sequences to be included in the delay test had to be presented at the first visit; addition of six more sequences would have posed a substantial burden). (Preliminary analyses revealed no difference in performance after 15 min for sequences tested at the first and second sessions: t test values < 1.55 , $ps > .10$, for actions and order, respectively.) The 15 min between presentation and test of the sequences used to assess the success of encoding was filled with presentation and/or test of the other sequences in this condition. That is, all three of the sequences were modeled in turn and then once 15 min had passed, they were tested in turn, in the same order in which they had been presented. Across participants, each sequence was used equally often in each encoding condition.

It should be noted that in this research, infants were not given a baseline opportunity for interaction with the props for each sequence prior to modeling of it. This represents a departure from the procedures typically employed in studies using elicited imitation (e.g., Bauer et al., 2000). This decision was made in order to accommodate the large number of sequences and learning trials necessary for a within-subjects test of the motivating hypotheses while keeping the test sessions within the range of tolerance of the young participants. As noted earlier, all of the test sequences had been used in prior studies with infants of comparable ages (e.g., Bauer & Mandler, 1989; Bauer et al., 1999, 2000; Wiebe & Bauer, 2005). In all cases, infants' performance after modeling was significantly higher than performance prior to modeling (i.e., in baseline). Thus, we are confident that infants' postmodeling performance can be attributed to their memory for the modeled sequences.

Scoring—At both sessions, infants' performance was coded by the experimenter during the laboratory visits or, in some cases, shortly thereafter from the videotapes. For each sequence, we calculated the total number of different target actions produced (maximum = 3) and the number of pairs of actions produced in the target order (maximum = 2). The first measure indicates the extent to which the infants remembered the individual elements of the sequences; the second measure indicates the extent to which they created temporally

organized representations of them. In calculating the number of pairs of actions produced in the target order, we considered only the first occurrence of each target action. For example, if infants produced the string of target actions 3-1-2-3, they would be credited with three different target actions, but with only one correctly ordered pairs of target actions, namely, 1-2. They would not be credited with the correctly ordered Pair 2-3, because they would have received credit for the first production of Target Action 3. This scoring procedure reduces the likelihood of credit for production of an ordered sequence by chance or by trial and error.

For purposes of reliability, 30 sessions (25% of the sample) were independently recoded from videotapes. Overall agreement between the experimenters/coders was 91% (range = 70–100%).

Results

Preliminary analyses—To accomplish the comparison of three encoding conditions (watch, imitate, criterion), it was necessary for infants to reach criterion on at least one of the two sequences that would be tested after the 15-min delay and at least one of the two sequences that would be tested 1, 2, or 3 weeks later. Although 60 infants participated in the experiment ($N = 20$ per delay condition), in the 1-, 2-, and 3-week delay conditions, only 15, 13, and 11 infants, respectively, met the requirement of reaching criterion on at least one sequence to be used in the 15-min test and at least one sequence to be used in the weeks test (details on the numbers of infants who reached criterion on 0, 1, or 2 sequences at each test, in each group, and the average numbers of trials to criterion at each test, in each group, are available from the first author). Whereas the different numbers of infants who reached criterion suggests that random assignment might have resulted in unequal groups, analysis of infants' immediate test performance in the imitate condition suggests otherwise. That is, one-way analyses of variance (ANOVAs) of the number of actions and number of pairs of actions in the correct temporal order produced immediately after modeling in the imitate condition revealed no significant between-groups differences, $F_s < 1.75$, $p_s > .19$. All subsequent analyses were based on the samples of infants who reached criterion. For the criterion condition, the measure of performance was the score on the one sequence at each test on which the infant had reached criterion, or the mean of the two sequences if the infant reached criterion on both. For the other encoding conditions (watch, imitate), the measure of performance was the mean of the two sequences per test/condition.

The integrity of the approach taken in this experiment rests on the assumption that all of the sequences within an encoding condition were learned equally well, and thus, that the two sequences tested at each delay (15 min; 1, 2, or 3 weeks) constituted a representative sample of the success of encoding and of consolidation, respectively. In the criterion condition, this assumption was met by virtue of the fact that only sequences on which infants reached criterion were included in the analyses. In the imitate condition, the assumption was tested by comparing immediate recall performance on the four sequences in the imitate condition. The one-way ANOVAs of the number of actions and number of pairs of actions in the correct temporal order produced immediately after modeling on the four sequences in the imitate condition revealed no significant differences, $F_s < 2.60$, $p_s > .05$. Although there

was no means of testing the assumption in the watch condition (i.e., no measures of immediate recall were taken), given that the condition was met in the criterion and imitate conditions, and given that sequences were counterbalanced across conditions, it is reasonable to assume that the assumption of equal learning was met in the watch condition as well.

Main analyses—Descriptive statistics on the levels of production of the individual actions of the sequences and the pairs of actions in the correct temporal order are provided in Table 2. To examine performance at the different test points as a function of delay group and encoding condition, we conducted separate 3 (delay group: 1, 2, 3 weeks) $\times$ 3 (encoding condition: watch, imitate, criterion) $\times$ 2 (test: 15 min, weeks) mixed ANOVAs, for the variables of actions and pairs produced, with delay group as a between-subjects variable. Because the numbers of infants included in the analyses differed across delay groups (i.e., because different numbers of infants reached criterion in the criterion encoding condition), we used *SAS Proc GLM* (Statistical Analysis Software, Cary, NC) procedure for the analyses. Tukey tests ($p < .05$) were used to examine main effects involving more than two means. The analyses revealed main effects of test, for both dependent measures, $F_s(1, 36) = 10.58$ and 16.88 , $p_s < .003$, $\eta^2 = .042$ and $.053$, for actions and order, respectively. For both dependent measures; performance after 15 min was higher than performance after the weeks delay. Thus, there was significant forgetting between the shorter (15 min) and longer (1–3 weeks) delay intervals. There was not a main effect of delay group, suggesting that performance in the 1-, 2-, and 3-week conditions was not substantially different.

For both dependent measures there were main effects of encoding condition. The main effects were qualified by interactions with test, $F_s(2, 72) = 5.54$ and 5.75 , $p_s < .006$, $\eta^2 = .035$ and $.036$, for actions and order, respectively. Follow-up analyses of the interactions revealed different patterns of recall at the 15-min and weeks tests. The pattern was the same for the two dependent measures. After 15 min, the number of actions and number of pairs of actions in correct temporal order that the infants produced did not differ across the encoding conditions. To quantify the amount of variance explained by encoding condition, we conducted regression analyses. The analyses revealed that after 15 min, the proportion of variance accounted for by encoding condition was a nonsignificant 1% ($p > .25$) and 2% ($p > .10$) for actions and pairs, respectively. At the weeks test, there were main effects for encoding condition, $F_s(2, 114) = 7.41$ and 14.44 , $p_s < .001$, $\eta^2 = .115$ and $.202$, for actions and pairs, respectively. Tukey tests ($p < .05$) revealed that for both dependent measures, after delays of 1–3 weeks, the infants produced fewer actions and pairs of actions in the target order in the watch encoding condition relative to the imitate and criterion encoding conditions, which did not differ from one another (see Table 2 for means and standard deviations). Regression analyses revealed that after 1–3 weeks, the proportion of variance accounted for by encoding condition was 7% and 13%, for actions and pairs, respectively, $F_s(2, 192) = 7.87$, $p < .0005$, and 15.73 , $p < .0001$, for actions and pairs, respectively. For illustrative purposes, the changes in levels of performance from the 15-min test to the weeks test over delays of 1–3 weeks, in the watch, imitate, and criterion encoding conditions, are depicted in Figure 1.

Finally, for both dependent variables there were Delay group $\times$ Encoding condition interactions, $F_s(4, 72) = 2.84$ and 2.58 , $ps < .05$, $\eta^2 = .041$ and $.035$, for actions and order, respectively. Follow-up tests revealed only one significant difference. In the watch encoding condition, the infants in the 1-week delay group produced fewer individual target actions than infants in the 2-week delay group, $F(2, 36) = 4.53$, $p < .02$, $\eta^2 = .090$ (Tukey $p < .05$). A similar trend was observed for the variable of pairs of actions in the target order. However, the effect was not statistically significant ($p = .10$). None of the other comparisons was significant. Given that the significant difference in production of individual target actions was isolated, its most likely source was chance variation among the groups.

Discussion

The purpose of this experiment was to investigate possible differential post-encoding processing as a function of encoding condition. We contrasted three encoding conditions, two of which had been shown to result in different levels of long-term recall (watch versus imitate) and the third of which ensured complete encoding, but which had not previously been directly compared with other encoding conditions.

Different relative levels of performance across the encoding conditions were observed 15 min after exposure to the sequences compared to 1, 2, or 3 weeks later. After 15 min, there were no differences in performance as a function of encoding condition. In contrast, 1–3 weeks later, infants had lower levels of recall of the individual target actions and the temporal order of actions of the sequences in the watch condition relative to the sequences in the imitate and criterion conditions. Because the manipulation of encoding condition was within-subjects, it is likely that the differences in recall after the weeks delays were because of differentially effective post-encoding processes, as opposed to differential retrieval success. That is, the only reason why retrieval cues would be less effective in the watch relative to the imitate and criterion conditions is that the memory traces were differentially intact at the time of retrieval. The pattern of findings is consistent with the suggestion of differential post-encoding processes that ultimately impact long-term recall.

We sound two cautions in interpretation of the findings from this experiment. First, findings from the criterion encoding condition were based on a smaller number of infants, relative to those in the imitate and watch conditions. The difference was because of elimination from the sample infants who did not reach criterion on at least one sequence to be used in the 15-min test and at least one sequence to be used in the weeks test. In the 1-, 2-, and 3-weeks delay groups, application of the criterion resulted in loss of five, seven, and nine infants, respectively. This level of exclusion is higher than that in the only other infant study in which a criterion encoding condition was used, namely, Howe and Courage (1997). Importantly, Howe and Courage administered up to 10 learning trials whereas in this research, we administered a maximum of four. Based on pilot research, we determined that the lower maximum number of trials was necessary to ensure infants' participation in all three conditions required for execution of the within-subjects design.

The second caution in interpretation of the results of this experiment is that the assumption of equal encoding across the conditions was based on a test of performance administered 15

min after exposure to the test sequences. We consider the 15-min test a valid proxy for the success of encoding, based on prior research with same-age infants in which there was no decrement in performance 10–15 min after modeling relative to immediate imitation conditions (i.e., Bauer et al., 1999; Cheatham et al., in press). Nevertheless, the fact that there was no direct measure of the success of encoding in the watch condition must be kept in mind in interpreting the results of this research.

By testing long-term recall after 1, 2, and 3 weeks, the design of Experiment 1 was maximally similar to research with animal models, in which the time course of systems-level consolidation has been tested (e.g., Kim & Fanselow, 1992; Takehara et al., 2003). The relatively coarse grain of the examination did not allow for identification of the timing of the change in relative levels of performance, however. That is, at the shortest weeks delay interval (1 week), the level of performance on sequences in the watch condition already had decreased, relative to the levels of performance in the imitate and criterion conditions. There was no further change in relative levels of performance with increasing delay. To gain a more fine-grained perspective on the time course of changes in relative levels of performance, in Experiment 2, we measured the success of consolidation over days rather than weeks. The experiment presents an opportunity for conceptual replication of Experiment 1, as well as test of when differential forgetting first becomes apparent.

Experiment 2: Minutes and Days

Method

Participants—The participants were 60 infants with a mean age of 19 months, 24 days (range = 19 months 14 days to 20 months 15 days). Half of the infants were girls. The infants were drawn from the same source and represent the same population as in Experiment 1. None of the infants had participated in the prior experiment. As in Experiment 1, the infants were pseudo-randomly assigned to delay conditions, with the provision that an equal number of infants was assigned to each delay condition ($N = 20$), and each delay condition included an equal number of girls and boys. At the end of the second session, all infants received a small toy for their participation.

Material, procedure, and scoring—The materials were the same 12 three-step sequences used in Experiment 1 (see Appendix). The procedure was the same, with the exception that the long delays were 1, 2, and 4 days after exposure to the test sequences (rather than 1, 2, and 3 weeks later). Scoring was conducted as in Experiment 1. Reliability between coders, calculated on 30 sessions (25% of the sample), was 92% (range = 71–100%).

Results

Preliminary analyses—As observed in Experiment 1, there was between-group variability in the number of infants in the criterion encoding condition who reached the criterion of two successive accurate reproductions of all of the actions of the sequences in order on at least one sequence to be used in the 15-min and at least one sequence to be used in the days test: in the 1, 2, and 4-day delay conditions, 15, 17, and 13 of the children

reached criterion, respectively, (details on the numbers of infants who reached criterion on 0, 1, or 2 sequences at each test, in each group, and the average numbers of trials to criterion at each test, in each group, are available from the first author). One-way ANOVAs on the number of actions and number of pairs of actions in the correct temporal order produced immediately after modeling in the imitate encoding condition revealed no significant between-groups differences), $F_s < 1.00$, $p_s > .40$. Thus, as in Experiment 1, there was no evidence that random assignment resulted in unequal groups. Also as in Experiment 1, in the imitate condition, one-way ANOVAs of the number of actions and number of pairs of actions in the correct temporal order produced immediately after modeling revealed no significant differences across the four sequences, $F_s < 1.00$, $p_s > .40$. Thus, there was not unequal learning of the sequences that would be used in the 15-min and days delay conditions. Finally, preliminary analyses revealed no difference in 15-min performance for sequences tested at the first and second sessions (t test values < 1.00 , $p_s > .50$).

Main analyses—Descriptive statistics on the levels of production of the individual actions of the sequences and the pairs of actions in the correct temporal order are provided in Table 3. To examine performance at the different test points as a function of delay group and encoding condition, using the *SAS Proc GLM* procedure, we conducted separate 3 (delay group: 1, 2, 4 days) $\times$ 3 (encoding condition: watch, imitate, criterion) $\times$ 2 (test: 15 min, days) mixed ANOVAs, for the variables of actions and pairs produced, with delay group as a between-subjects variable. Neither analysis revealed main effects of delay group or test. Thus, overall, there was no evidence of differential production of either the individual actions or the temporal order of actions by infants in the 1- versus the 2- versus the 4-day delay groups. The absence of main effects of test indicates that overall, there were no substantial differences in performance between tests. For the variable of pairs of actions in the correct temporal order, the interaction of delay group and test was significant, $F(2, 42) = 3.54$, $p < .05$, $\eta^2 = .017$. However, follow-up analyses revealed no significant differences either as a function of group or as a function of delay.

For both dependent measures there were main effects of encoding condition. The main effects were qualified by interactions with test, $F_s(2, 83) = 10.93$ and 10.28 , $p_s < .0001$, $\eta^2 = .059$ and $.049$, for actions and temporal order, respectively. Follow-up analyses of the interactions revealed different patterns of recall of the actions of the sequences and their order, and different patterns at the 15-min and days tests. In terms of the number of individual actions produced, there were main effects of encoding condition at both tests, $F_s(2, 132) = 3.91$ and 34.40 , $p_s < .03$ and $.0001$, $\eta^2 = .056$ and $.344$, for 15-min and days, respectively. Tukey tests ($p < .05$) revealed that after 15 min, the infants produced more actions in the criterion relative to the imitate encoding condition. Production of actions in the watch encoding condition was intermediate and did not differ from either the criterion or imitate condition (see Table 3 for means and standard deviations). Regression analyses revealed that after 15 min, the proportion of variance accounted for by encoding condition was a nonsignificant 3% ($p > .20$). After 1–4 days, the infants recalled fewer actions in the watch encoding condition, relative to the criterion and imitate encoding conditions, which did not differ from one another. Regression analyses revealed that after 1–4 days, the proportion of variance accounted for by encoding condition was 20%, $F(2, 220) = 29.36$, p

< .0001. For illustrative purposes, the changes in levels of performance from the 15-min test to the days test over delays of 1–4 days, in the watch, imitate, and criterion encoding conditions, are depicted in Figure 2. As depicted in the figure, the change in relative levels of performance was because of a nominal decrease in production of the actions of the sequences in the watch condition, in the face of nominal increases in the imitate and criterion conditions.

In terms of production of pairs of actions in the correct temporal order, there were main effects of encoding condition at both tests, $F_s(2, 132) = 6.59$ and 32.48 , $p_s < .002$ and $.0001$, $\eta^2 = .091$ and $.331$, for 15-min and days, respectively. Tukey tests ($p < .05$) revealed that after 15 min, the infants produced more pairs of actions in the correct order in the criterion relative to the imitate and watch encoding conditions, which did not differ from one another. Regression analyses revealed that after 15 min, the proportion of variance accounted for by encoding condition was 6%, $F(2, 220) = 17.70$, $p < .0001$. After 1–4 days, the infants produced significantly more pairs of actions in the correct temporal order in the criterion relative to the imitate encoding condition. In turn, they produced significantly more pairs of actions in the correct temporal order in the imitate than in the watch encoding condition. Regression analyses revealed that after 1–4 days, the proportion of variance accounted for by encoding condition was 20%, $F(2, 220) = 31.58$, $p < .0001$. As depicted in Figure 2, the change in relative levels of performance was because of a nominal decrease in production of ordered pairs of actions of the sequences in the watch condition, in the face of nominal increases in the imitate and criterion conditions.

Discussion

Fifteen minutes after exposure, infants had higher levels of production of the individual actions and temporal order of actions of the sequences in the criterion relative to the imitate condition. They also had higher levels of production of the temporal order of actions of the sequences in the criterion relative to the watch condition. In contrast, days after exposure, production of the individual actions of the sequences no longer differed in the criterion and imitate conditions, though the difference in ordered recall remained. Performance in the watch condition was significantly lower than that in the criterion and imitate conditions, on both dependent measures. Thus, the actions of sequences that initially had not been as well encoded (imitate relative to criterion) were better recalled days later, whereas the actions of sequences that initially had been well encoded (watch relative to criterion) were less well recalled. The pattern of findings is consistent with the suggestion of differential post-encoding processes that ultimately impact long-term recall.

By testing long-term recall after 1, 2, and 4 days, the design of Experiment 2 was potentially sensitive to detection of significant forgetting as well as the point in time at which it might set in. This feature of the design was not realized, however: no significant forgetting was detected across the 1-, 2-, or 4-day delays.

General Discussion

The purpose of this research was to investigate possible differential post-encoding processing as a function of encoding condition. Twenty-month-old infants were exposed to three-step event sequences under two different encoding conditions that in previous research have been associated with differential levels of long-term recall, namely, sequences they were permitted to imitate immediately after demonstration and sequences for which imitation was deferred. In addition, they experienced a criterion encoding condition that ensured full encoding and thus was expected to be associated with high levels of delayed recall. Performance in the three encoding conditions was compared 15 min after experience of the events and again either weeks later (Experiment 1) or days later (Experiment 2). The tests provided indices of the efficacy of encoding of the events and consolidation of the memory traces, respectively.

Because the conditions for testing 15 min after experience of the events were the same in the two experiments, we can ask whether patterns of performance were the same. As summarized in Table 4, there were some differences in performance at the 15-min tests in Experiments 1 and 2. In Experiment 1, no significant differences in levels of performance across encoding conditions were detected, whereas in Experiment 2, performance in the criterion condition was greater than that in the watch condition (pairs of actions only) and in the imitate condition (actions and pairs). The differences are attributable to (a) nominally lower levels of performance in the watch and imitate conditions in Experiment 2 relative to Experiment 1, with comparable levels of performance in the criterion condition (see Tables 2 and 3), and (b) coupled with greater power to detect differences in Experiment 2 relative to Experiment 1, as a result of the larger number of infants who met criterion in the second relative to the first experiment (45 and 39 infants, respectively).

In spite of the fact that long-term recall was tested weeks versus days later in Experiments 1 and 2, respectively, there was only one difference in patterns of performance in the two experiments. As reflected in Table 4, whether long-term recall was tested weeks or days later, performance in both the criterion and imitate conditions was greater than performance in the watch conditions. One difference between the experiments was in comparison of performance in the imitate versus criterion condition. Whereas in Experiment 1, the conditions did not differ, in Experiment 2, infants produced a larger number of ordered pairs in the criterion relative to the imitate condition. Because the size of the difference in levels of performance was roughly the same in the two experiments (.23 and .26, in Experiments 1 and 2, respectively), the most likely source of the difference in levels of statistical significance is the greater power in Experiment 2 relative to Experiment 1.

For the purposes of this research, more important than qualitative comparisons between experiments are quantitative, within-experiment comparisons of the 15-min and later tests. In both experiments, different patterns of performance were observed 15 min after experience of the events and weeks (1, 2, or 3) or days (1, 2, or 4) later. Specifically, as summarized in Table 4, whereas 15 min after experience of the events, performance on sequences in the watch encoding condition largely was on par with performance on events in the criterion condition, weeks later and days later, performance in the watch condition was

significantly lower. Similarly, whereas after 15 min, performance in the watch and imitate encoding conditions did not differ, weeks later and days later, performance in the watch condition was significantly lower. Relative levels of performance on sequences in the imitate and criterion conditions remained relatively stable both weeks later and days later. Consistent with the suggestion of differentially successful consolidation of events in the face of equal learning of them, encoding condition accounted for little variance in performance after only 15 min, but for as much as 20% of variance in performance after weeks or days. Thus, knowing how a sequence was experienced was not especially predictive of how successfully it was learned (as measured by performance after 15 min), but it was predictive of how successfully it was consolidated.

What accounts for differentially successful consolidation in the face of equal learning? As reviewed earlier, consolidation of new memory traces is considered to involve both the stabilization of the trace through formation of associations among the individual elements of experience, and integration of the trace into cortical association areas (e.g., Zola & Squire, 2000). It is reasonable to expect that both of these processes would be facilitated by a more broadly distributed representation that involves multiple modalities and thus, as argued by Lukowski et al. (2005), multiple brain regions. The experience involved in production of event sequences in the imitate and criterion conditions ensured motor inputs to the representation, as well as visual and auditory inputs (stemming from infants' actions on the props, the props themselves, and the narration provided by the experimenter, respectively). In contrast, watching event sequences being modeled engaged visual and auditory regions, but entailed less involvement of motor regions. The implications of a greater number of inputs to the memory trace would grow over time as the individual elements of the representation were associated with one another, and as the elements of the representation were integrated with traces already in storage. In both cases, more elements would mean more opportunities for association, thereby maximizing the likelihood that the trace would survive the period of consolidation.

The importance of post-encoding processes has been apparent since the beginning of the 20th century (Müller & Pilzecker, 1900). Yet in both the adult cognitive (Wixted, 2004) and the developmental (Bauer, 2006, 2009b) literatures, they have been relatively neglected. In the developmental literature, rather than on post-encoding processes, most of the work concerned with age-related differences in long-term recall has implicated either encoding or retrieval as the major source of variance. This research makes clear that even in the face of equal encoding, systematic variability in subsequent recall is apparent. The fact that the manipulation of encoding condition was carried out within, rather than between-subjects, and that the retrieval test was the same for all encoding conditions, diminishes the plausibility of retrieval processes as the major source of observed variability. With the “book ends” of encoding and retrieval accounted for, the most logical source of variance is in processes that take place after encoding, but prior to retrieval.

The results of this research are consistent with the small body of work that permits examination of post-encoding processes as a source of variance in long-term recall. In Bauer (2005), infants 13 and 16 (Experiment 1) and 16 and 20 (Experiment 2) months of age were matched for encoding yet exhibited differential levels of recall 1–6 months later. In Howe

and Courage (1997), infants 15 and 18 (Experiment 1) and 12 and 15 (Experiment 2) months of age reached a criterion level of learning—and thus complete encoding—yet exhibited differential levels of recall 3 months later. This research adds to this literature by examining the question within rather than between-subjects, and by testing long-term recall after various timepoints days and weeks after encoding.

As already noted, the within-subjects manipulation has the advantage of virtually eliminating an alternative explanation for differential long-term recall, namely, differential retrieval success. Another advantage of this approach is that it afforded comparison of learning and long-term recall across three different encoding conditions. Although the criterion encoding condition was expected to produce the highest levels of performance, after 15 min it did not differ from performance in the watch condition, and only in Experiment 2 did it differ from performance in the imitate condition. Days and weeks later, the criterion encoding condition offered clear advantages relative to the watch condition. Relative to the imitate condition, the only advantage afforded by criterion encoding was in ordered recall tested after 1–4 days. Relative to the watch condition, the imitate condition afforded clear advantage days and weeks later. There was no advantage after only 15 min had passed.

By testing recall at several different times days and weeks after experience of events, this research allowed some insight into the time course of forgetting among infants in the second half of the second year of life. Contrary to observations in prior related research (e.g., Bauer et al., 1999; Hanna & Meltzoff, 1993; Tomasello et al., 1993), forgetting was not apparent even 96 hrs (4 days) after experience of events (Experiment 2). Although forgetting had set in by the time 1 week had elapsed (Experiment 1), no additional forgetting between 1 and 3 weeks was detected. These findings attest to the robustness of event memory among 20-month-old infants, under a variety of encoding conditions. They indicate that even though such young infants may suffer substantial loss of information during the period of consolidation, much remains to be carried forward to another day (or week).

This research was an important step in evaluating post-encoding processes as a potential source of variance in long-term recall. Two next steps are especially salient. First, it will be important to examine the implications of post-encoding processes assessed during the period of consolidation for recall of events over the long term. Although the time course of consolidation is not clear, the animal literature suggests that once 1 month has passed since the experience of an event, the resulting memory trace may be considered to be in storage as opposed to still undergoing active consolidation (e.g., Kim & Fanselow, 1992; Takehara et al., 2003). An important question is whether samples of the “health” of memory traces during the period of consolidation explain significant variance in the integrity of traces over retention intervals that extend beyond the period. Suggestive evidence that they do was provided in Bauer, Wiebe, Carver, Waters, and Nelson (2003). They found that event-related potential indices of the health of memory traces 1 week after experience of events explained 28% of the variance in 9-month-olds' recall 1 month later. Similarly, in Bauer, Cheatham, Cary, and Van Abbema (2002), measures of performance 48 hrs after exposure to events explained 25% of the variance in 20-month-olds' recall after 1 month. Additional and more

systematic investigations of this sort will permit address of the question of relations between post-encoding processes and long-term recall.

The literature also will be advanced by examinations of possible age-related differences in post-encoding processes. Consolidation processes were suggested as a potential source of age-related variability in long-term recall because of the relative immaturity of the neural structures and network implicated in them (Bauer, 2006). Logically, as the substrate develops, age-related variability in post-encoding processes should diminish (though individual differences will remain). Test of this suggestion will require comparison of performance at different delay intervals, at multiple ages, and comparison of the variance in long-term recall explained at each age.

In conclusion, this research permitted test of the possibility of differences in post-encoding processes as a function of differences in the conditions of encoding. We found that whereas encoding condition explained little variance minutes after experience of events, it explained significant variance after delays of weeks (Experiment 1) and days (Experiment 2). The research highlights the importance of considering post-encoding processes as a source of variance in long-term recall. It paves the way for future research that will test the relative amounts of variance explained by probes of the integrity of memory traces at different points in their construction and maintenance.

Acknowledgments

Support for this research was provided by the NICHD (HD28425 and HD42483) and Emory University. The authors also thank Melissa Burch and the infants and families that so generously gave of their time to make the work possible. Portions of these data were presented at the meetings of the Cognitive Development Society, Park City, Utah, 2003, and the International Society for Infant Development, Chicago, Illinois, 2004.

Appendix

Included here are descriptions of the 12 event sequences used in Experiments 1 and 2. All of the event sequences involved three steps. The props given to the children are indicated in parentheses. In each experiment, each event sequence was used approximately equally often in each delay condition.

1. *Gong* (base resembling the support for a swing set with the top bar movable on a hinge; “bell” made of a square of metal with a curved edge to hang it; hammer). Fold the bar across the support in order to form a cross piece; hang the bell from the cross piece; “ring” the bell with the hammer.
2. *Paddle rattle* (paddle with a handle; block; stacking box). Put the block on the paddle; cover it with the box; shake the handle to make it rattle.
3. *Spinner* (circular plastic base with a raised peg in the center; center piece with plastic horses that fit over the raised peg; dome-shaped cover with a depressible “button” protruding from the top). Put the horse center piece on the base; cover the base and center piece with the dome; push the button to make the horses spin.
4. *Shaker* (two nesting cups; small block). Put the block inside one cup; cover it with the other cup; shake the cups.

5. *Fruit chute* (ramp piece with a hinged middle; base that functioned as an incline; plastic plum). Open the ramp to its full length; place the ramp on the base to form an incline; roll the plum down the ramp.
6. *Horse ride* (rectangular base with a plunger on one end and a hole in the top; plastic stick; horse body with hole in the center). Put the stick into the hole at the top of the base; put the horse onto the stick; push the plunger in and out to make the horse move back and forth on the base.
7. *Merry-Go-Round* (circular base with a wooden rod extending vertically from the center; semicircular hinged top piece that when closed formed a circular piece with a hole in the middle; three wooden horses hung from the edge of the top). Close the top piece; put the top onto the base; spin the top to make the horses go around.
8. *Wheel* (horizontal wooden base with one fixed vertical side and one hinged side; circular piece of wood with short pegs protruding from both sides). Raise the hinged side of the base; place the wheel on the base; hit the wheel to make it spin.
9. *Drum* (round acrylic base with a plastic notch on the top rim; L-shaped metal rod; circular metal cover). Place the rod in the notch; put on the cover; press the rod to “drum it.”
10. *Motion drum* (base with stick protruding from the center; solid ring with hole in the center; barber-pole cylinder). Place the ring on the base; position the cylinder over the ring; spin the cylinder to make a barber-pole effect.
11. *Jumper* (base with circular indentation and handle protruding from the side; clear globe with ball inside). Place the globe on the base; push the ball out of the globe onto the base; press the handle to make the ball “jump” up and down on the base.
12. *Dancing bear* (stick with Velcro at one end; circular base with a hole in the center; bear with movable limbs, pull-string, and Velcro on back). Insert the non-Velcro end of the stick into the base; fasten the bear to the stick with Velcro; pull the bear's string, resulting in a “dancing” motion of its limbs.

References

- Barr R, Dowden A, Hayne H. Developmental change in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development*. 1996; 19:159–170.
- Barr R, Hayne H. The effect of event structure on imitation in infancy: Practice makes perfect? *Infant Behavior and Development*. 1996; 19:253–257.
- Bauer PJ. Holding it all together: How enabling relations facilitate young children's event recall. *Cognitive Development*. 1992; 7:1–28.
- Bauer, PJ. Development of memory in early childhood. In: Cowan, N., editor. *The development of memory in childhood*. Sussex, UK: Psychology Press; 1997. p. 83–111.
- 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. 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. Learning and memory: Like a horse and carriage. In: Woodward, A.; Needham, A., editors. *Learning and the infant mind*. New York: Oxford University Press; 2009a. p. 3-28.
- 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; 2009b. p. 78-102.
- 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. 9. Huntington, NY: Nova Science Publishers; 2002. p. 53-74.
- Bauer PJ, Hertsgaard LA, Dropik P, Daly BP. When even arbitrary order becomes important: Developments in reliable temporal sequencing of arbitrarily ordered events. *Memory*. 1998; 6:165–198. [PubMed: 9640427]
- Bauer PJ, Hertsgaard LA, Wewerka SS. Effects of experience and reminding on long-term recall in infancy: Remembering not to forget. *Journal of Experimental Child Psychology*. 1995; 59:260–298. [PubMed: 7722436]
- Bauer PJ, Mandler JM. One thing follows another: Effects of temporal structure on one- to two-year-olds' recall of events. *Developmental Psychology*. 1989; 25:197–206.
- Bauer PJ, Thal DJ. Scripts or scraps: Reconsidering the development of sequential understanding. *Journal of Experimental Child Psychology*. 1990; 50:287–304. [PubMed: 2258692]
- Bauer PJ, Van Abbema DL, de Haan M. In for the short haul: Immediate and short-term remembering and forgetting by 20-month-old children. *Infant Behavior and Development*. 1999; 22:321–343.
- 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(4, Serial No. 263)
- 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]
- Bourgeois, JP. Synaptogenesis in the neocortex of the newborn: The ultimate frontier for individuation?. In: Nelson, CA.; Luciana, M., editors. *Handbook of developmental cognitive neuroscience*. Cambridge, MA: The MIT Press; 2001. p. 23-34.
- Brown AS. Consolidation theory and retrograde amnesia in humans. *Psychonomic Bulletin and Review*. 2002; 9:403–425. [PubMed: 12412884]
- Burch, MM.; Schwade, JA.; Bauer, PJ. Finding the right fit: Examining developmentally appropriate levels of challenge in elicited imitation studies. In: Bauer, PJ., editor. *Varieties of early experience: Implications for the development of declarative memory in infancy Advances in child development and behavior*. Vol. 38. London: Elsevier; in press
- Cheatham, CL.; Sesma, HW.; Bauer, PJ.; Georgieff, MK. The development of declarative memory in infants born preterm. In: Bauer, PJ., editor. *Varieties of early experience: Implications for the development of declarative memory in infancy Advances in child development and behavior*. Vol. 38. London: Elsevier; in press
- Eichenbaum, H.; Cohen, NJ. *From conditioning to conscious recollection: Memory systems of the brain*. New York: Oxford University Press; 2001.
- Gluck, MA.; Myers, CE. *Gateway to memory: An introduction to neural network modeling of the hippocampus and learning*. Cambridge, MA: The MIT Press; 2001.
- Goldman-Rakic, PS. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In: Plum, F., editor. *Handbook of physiology, the nervous system, higher functions of the brain*. Vol. 5. New York: American Physiological Society; 1987. p. 373-417.
- Hanna E, Meltzoff AN. Peer imitation by toddlers in laboratory, home, and day-care contexts: Implications for social learning and memory. *Developmental Psychology*. 1993; 29:702–710.
- Hayne H, Barr R, Herbert J. The effect of prior practice on memory reactivation and generalization. *Child Development*. 2003; 74:1615–1627. [PubMed: 14669885]
- Hayne H, Boniface J, Barr R. The development of declarative memory in human infants: Age-related change in deferred imitation. *Behavioral Neuroscience*. 2000; 114:77–83. [PubMed: 10718263]

- 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. *Journal of Comparative Neurology*. 1997; 387:167–178. [PubMed: 9336221]
- Kandel ER, Squire LR. Neuroscience: Breaking down scientific barriers to the study of brain and mind. *Science*. 2000; 290:1113–1120. [PubMed: 11185010]
- Kim JJ, Fanselow MS. Modality-specific retrograde amnesia of fear. *Science*. 1992; 256:675–677. [PubMed: 1585183]
- Klein PJ, Meltzoff AN. Long-term memory, forgetting and deferred imitation in 12-month-old infants. *Developmental Science*. 1999; 2:102–113. [PubMed: 25147475]
- Liston C, Kagan J. Memory enhancement in early childhood. *Nature*. 2002; 419:896. [PubMed: 12410299]
- Lukowski AF, Wiebe SA, Haight JC, DeBoer T, Nelson CA, Bauer PJ. Forming a stable memory representation in the first year of life: Why imitation is more than child's play. *Developmental Science*. 2005; 8:279–298. [PubMed: 15819758]
- McGaugh JL. Memory—A century of consolidation. *Science*. 2000; 287:248–251. [PubMed: 10634773]
- Meltzoff AN. Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development*. 1985; 56:62–72.
- Meltzoff AN. Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology*. 1988a; 24:470–476. [PubMed: 25147404]
- Meltzoff AN. Imitation of televised models by infants. *Child Development*. 1988b; 59:1221–1229. [PubMed: 3168638]
- Müller GE, Pilzecker A. Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift für Psychologie*. 1900; 1:1–300.
- Nelson, CA. The neurobiological basis of early memory development. In: Cowan, N., editor. *The development of memory in childhood*. Hove East Sussex: Psychology Press; 1997. p. 41–82.
- Reed JM, Squire LR. Retrograde amnesia for facts and events: Findings from four new cases. *The Journal of Neuroscience*. 1998; 18:3943–3954. [PubMed: 9570821]
- Schneider JFL, Il'yasov KA, Hennig J, Martin E. Fast quantitative diffusion-tensor imaging of cerebral white matter from the neonatal period to adolescence. *Neuroradiology*. 2004; 46:258–266. [PubMed: 14999435]
- 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*. 2nd. 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. *The Journal of Neuroscience*. 2003; 23:9897–9905. [PubMed: 14586019]
- Tomasello M, Savage-Rumbaugh S, Kruger AC. Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Development*. 1993; 64:1688–1705. [PubMed: 8112113]
- Wenner JA, Bauer PJ. Bringing order to the arbitrary: One- to two-year-olds' recall of event sequences. *Infant Behavior and Development*. 1999; 22:585–590.
- 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; 6:325–363.
- 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, E.; Craik, FIM., editors. *The Oxford handbook of memory*. New York: Oxford University Press; 2000. p. 485–500.

Figure 1.

Experiment 1—The change in levels of production of actions and correctly ordered pairs of actions from the 15-min test to the weeks test over delays of 1–3 weeks, in the watch, imitate, and criterion encoding conditions. The figure is for illustrative purposes only: analyses were conducted on raw scores (see Table 2), not the difference scores depicted in the figure.

Figure 2.

Experiment 2—The change in levels of production of actions and correctly ordered pairs of actions from the 15-min test to the weeks test over delays of 1–4 days, in the watch, imitate, and criterion encoding conditions. The figure is for illustrative purposes only: analyses were conducted on raw scores (see Table 3), not the difference scores depicted in the figure.

Table 1
Schematic Representation of Encoding Conditions (Panel a) and Testing Schedule (Panel b) for Experiment 1

Panel a: Representation of Encoding Conditions						
Encoding Condition	Modeling		Opportunity for Immediate Reproduction			
Watch	Sequence modeled 2 times		No			
	Sequence modeled 2 times		Yes			
Imitate	Sequence modeled 2 times		Yes			
	Sequence modeled 1 time		Yes			
Criterion (sequences modeled until 2 errorless imitations, or max 5 modelings, whichever came first)	Sequence modeled 1 time (as necessary)		Yes			
	Sequence modeled 1 time (as necessary)		Yes			
Panel b: Representation of Testing Schedule						
Encoding Condition	Mnemonic Process Assessed		Session 1		Session 2 (1, 2, or 3 weeks later)	
			Exposure	Test	Exposure	Test
Watch	Encoding		A	A		
	Encoding				J	J
Imitate	Consolidation		B			B
	Consolidation		C			C
	Encoding		D	D		
	Encoding				K	K
	Consolidation		E			E
	Consolidation		F			F
Criterion	Encoding		G	G		
	Encoding				L	L
	Consolidation		H			H
	Consolidation		I			I

Table 2
Experiment 1 (Weeks): Means (and Standard Deviations) for 20-Month-Olds' Performance as a Function of Delay Group, Encoding Condition, and Test

Encoding Condition					
Dependent Measure/Delay Group/Test	Watch	Imitate		Overall	
		M (SD)	M (SD)	M (SD)	M (SD)
Actions					
1 week	15 min	2.32 (.88)	2.18 (.78)	2.91 (.30)	2.45 (.68)
	1 week	1.64 (.74)	2.36 (.67)	2.59 (.66)	2.20 (.79)
2 weeks	15 min	2.63 (.35)	2.57 (.50)	2.63 (.90)	2.61 (.61)
	2 weeks	2.27 (.53)	2.40 (.76)	2.23 (.82)	2.30 (.70)
3 weeks	15 min	2.73 (.33)	2.35 (.63)	2.58 (.95)	2.55 (.69)
	3 weeks	1.85 (.52)	2.23 (.56)	2.77 (.39)	2.28 (.62)
Across groups	15 min	2.50 (.48)	2.38 (.63)	2.69 (.79)	2.55 (.65)
	Weeks	1.95 (.64)	2.33 (.66)	2.51 (.68)	2.26 (.70)
Pairs					
1 week	15 min	1.27 (.68)	1.14 (.71)	1.81 (.34)	1.41 (.65)
	1 week	.64 (.45)	1.36 (.50)	1.55 (.65)	1.18 (.66)
2 weeks	15 min	1.50 (.42)	1.57 (.42)	1.60 (.74)	1.56 (.54)
	2 weeks	1.00 (.38)	1.33 (.62)	1.23 (.75)	1.19 (.61)
3 weeks	15 min	1.50 (.41)	1.31 (.60)	1.58 (.73)	1.46 (.59)
	3 weeks	.81 (.38)	1.08 (.49)	1.73 (.39)	1.21 (.57)
Across groups	15 min	1.44 (.50)	1.36 (.58)	1.65 (.64)	1.48 (.59)
	Weeks	.83 (.42)	1.26 (.55)	1.49 (.64)	1.19 (.60)

Table 3
Experiment 2 (Days): Means (and Standard Deviations) for 20-Month-Olds' Performance as a Function of Delay Group, Encoding Condition, and Test

Dependent Measure/Delay Group/Test	Encoding Condition			
	Watch		Imitate	
	M (SD)	M (SD)	M (SD)	M (SD)
Actions				
1 day	15 min	2.50 (.57)	2.37 (.67)	2.73 (.59)
	1 day	2.03 (.44)	2.47 (.51)	2.97 (.13)
2 days	15 min	2.29 (.47)	2.06 (.88)	2.53 (.86)
	2 days	1.94 (.63)	2.76 (.50)	2.74 (.56)
4 days	15 min	2.54 (.63)	2.58 (.73)	2.88 (.42)
	4 days	2.13 (.68)	2.73 (.48)	2.92 (.28)
Across groups	15 min	2.43 (.55)	2.31 (.79)	2.70 (.67)
	Days	2.02 (.58)	2.66 (.51)	2.87 (.39)
Pairs				
1 day	15 min	1.33 (.49)	1.30 (.49)	1.67 (.62)
	1 day	.83 (.41)	1.33 (.59)	1.90 (.28)
2 days	15 min	1.12 (.57)	1.06 (.66)	1.53 (.67)
	2 days	.97 (.48)	1.67 (.60)	1.65 (.68)
4 days	15 min	1.42 (.67)	1.42 (.53)	1.77 (.39)
	4 days	.96 (.62)	1.46 (.43)	1.73 (.53)
Across groups	15 min	1.28 (.58)	1.24 (.58)	1.64 (.58)
	Days	.92 (.49)	1.50 (.52)	1.76 (.48)

Table 4
Summary of Relations Between Pairs of Encoding Conditions in Experiments 1 (Weeks)
and 2 (Days)

Pair of Conditions/Experiment/Dependent Measure		Delay Test	
		15 Min	Weeks/Days
Watch versus Imitate			
Experiment 1	Actions	NS	W < I
	Pairs	NS	W < I
Experiment 2	Actions	NS	W < I
	Pairs	NS	W < I
Watch versus Criterion			
Experiment 1	Actions	NS	W < C
	Pairs	NS	W < C
Experiment 2	Actions	NS	W < C
	Pairs	W < C	W < C
Imitate versus Criterion			
Experiment 1	Actions	NS	NS
	Pairs	NS	NS
Experiment 2	Actions	I < C	NS
	Pairs	I < C	I < C

Note. NS = nonsignificant; W = watch; I = imitate; C = criterion.