Received: 1 February 2016 | Accepted: 8 July 2016
DOI 10.1002/dev.21450

RESEARCH ARTICLE

WILEY Developmental Psychobiology

The impact of memory load and perceptual cues on puzzle learning by 24-month olds

Rachel Barr[1]* | **Alecia Moser[2]** | **Sylvia Rusnak[1]** | **Laura Zimmermann[1]** | **Kelly Dickerson[3]** | **Herietta Lee[1]** | **Peter Gerhardstein[2]**

[1] Department of Psychology, Georgetown University, Washington, District of Columbia

[2] Department of Psychology, Binghamton University-SUNY, Binghamton, New York

[3] Human Research and Engineering Directorate, United States Army Research Laboratory, Aberdeen Proving Ground, Maryland

***Correspondence**
Rachel Barr, Department of Psychology, Georgetown University, 3700 O St NW, Washington, DC 20057.
Email: rfb5@georgetown.edu

Funding Information
This research was supported by NSF (BSC-1023772).

Abstract

Early childhood is characterized by memory capacity limitations and rapid perceptual and motor development [Rovee-Collier (1996). *Infant Behavior & Development*, 19, 385–400]. The present study examined 2-year olds' reproduction of a sliding action to complete an abstract fish puzzle under different levels of memory load and perceptual feature support. Experimental groups were compared to baseline controls to assess spontaneous rates of production of the target actions; baseline production was low across all experiments. Memory load was manipulated in Exp. 1 by adding pieces to the puzzle, increasing sequence length from 2 to 3 items, and to 3 items plus a distractor. Although memory load did not influence how toddlers learned to manipulate the puzzle pieces, it did influence toddlers' achievement of the goal—constructing the fish. Overall, girls were better at constructing the puzzle than boys. In Exp. 2, the perceptual features of the puzzle were altered by changing shape boundaries to create a two-piece horizontally cut puzzle (displaying bilateral symmetry), and by adding a semantically supportive context to the vertically cut puzzle (iconic). Toddlers were able to achieve the goal of building the fish equally well across the 2-item puzzle types (bilateral symmetry, vertical, iconic), but how they learned to manipulate the puzzle pieces varied as a function of the perceptual features. Here, as in Exp. 1, girls showed a different pattern of performance from the boys. This study demonstrates that changes in memory capacity and perceptual processing influence both goal-directed imitation learning and motoric performance.

KEYWORDS
imitation, toddlers, learning, memory, sex differences

1 | INTRODUCTION

Early memory development is constrained by a number of factors, including changes in learning rate and duration of retention, as well as changes in memory flexibility (Barr & Hayne, 2000; Hayne, Rovee-Collier, & Perris, 1987; Rovee-Collier, Hayne, & Colombo, 2001). Although constrained by these basic processing limitations, early memory development also shares a number of core principles with the adult memory system (Rovee-Collier, 1997). Examination of these basic memory processes early in development, however, requires the adaptation of memory testing protocols to be developmentally appropriate for very young children. Carolyn Rovee-Collier pioneered operant conditioning paradigms for infants, including the mobile conjugate reinforcement paradigm (Rovee & Rovee, 1969) and the train paradigm (Hartshorn et al., 1998). Using these protocols, Carolyn and her colleagues demonstrated that, like adults, infants remember longer after more training sessions (Ohr, Fagen, Rovee-Collier, Hayne, & Vander Linde, 1989), and they remember more after longer training sessions (Ohr et al., 1989). Infants also show superior learning after spaced as compared to massed trials (Vlach, Sandhofer, & Kornell, 2008) and, in addition, they show the prior learning effect, where previously learned material is more easily recalled after new learning (Barr, Rovee-Collier, & Learmonth, 2011).

Early development is also characterized by rapid changes in motor and perceptual development. Simultaneous and significant changes in multiple processing systems means that increasing task complexity by increasing memory or perceptual load influences recall. Reducing the number of items to be remembered has been shown to improve recall by reducing cognitive load in adults (Baddeley, Thomson, & Buchanan, 1975; Cowan, 2001; Murdock, 1962; Shiffrin, 1993) and infants (Gulya, Sweeney, & Rovee-Collier, 1999; Gulya, Rovee-Collier, Galluccio, & Wilk, 1998; Merriman, Rovee-Collier, & Wilk, 1997).

Imitation paradigms, thought to tap declarative memory, provide an alternative non-verbal memory protocol for young children that researchers use to manipulate cognitive load (Hayne, 2006, for review). In these studies, experimenters demonstrate a series of actions on an object. Following a delay ranging from a few seconds to several weeks, infants are given an opportunity to copy what was observed during the demonstration (Barr, Dowden, & Hayne, 1996). The number of actions that infants reproduce at test indexes *deferred imitation*. Performance by the experimental groups is compared to that of a baseline control group not shown the demonstration of the target actions. Recall is inferred if the deferred imitation score of the experimental group exceeds that of the baseline control group (Barr & Hayne, 2000). Both infants (Bauer & Mandler, 1992; Kressley-Mba, Lurg, & Knopf, 2005) and preschoolers (Subiaul & Schilder, 2014) imitate shorter sequences with higher fidelity than longer sequences. On the contrary, increasing item number by using distractors—items that are irrelevant to achieving the goal—can increase cognitive load in 2-year olds (Wiebe & Bauer, 2005). These findings suggest that infants and children, like adults, are influenced by memory load task demands.

There are three ways by which children can accomplish a goal through imitation. These social learning strategies—action emulation, goal emulation, and imitation—have been observed across infancy and early childhood (Flynn & Whiten, 2013; Huang & Charman, 2005; Jones, 2009; Nielsen, 2006; Tennie, Greve, Gretscher, & Call, 2010; Want & Harris, 2002; Whiten, McGuigan, Marshall-Pescini, & Hopper, 2009). *Action emulation* (sometimes also referred to as mimicry) is defined as copying actions in the absence of goal-directed behavior (the how-to); *goal emulation* is defined as copying observed goals using any means available, including unobserved actions and those observed during the demonstration; while *imitation* is the combination of action and goal emulation—it occurs when both actions and goals are copied. Goal emulation is thought to be dependent on the ability to form abstract goal representations (Meltzoff, Waismeyer, & Gopnik, 2012). Some consider the use of different strategies to be developmentally tied, often citing action emulation as the earliest developing social learning strategy, with imitation and goal emulation to follow (Jones, 2007, 2009; Moser, Gerhardstein, Zimmermann, Grenell, & Barr, 2015; Tennie, Call, & Tomasello, 2006; Want & Harris, 2002). Want and Harris (2002) predicted that children would shift from action emulation strategies to a higher fidelity imitation strategy as a function of age, but Nielsen (2006) predicted that strategy shift should vary as a function of task demands (see also Over & Carpenter, 2012).

Despite the potential for deferred imitation tasks to tap into memory and social learning strategies, many deferred imitation tasks only assess goal-directed behaviors and restrict the range of producible behaviors by selecting objects-for-action that have a very limited number of affordances (Barnat, Klein, & Meltzoff, 1996; Hayne, Boniface, & Barr, 2000; see Hayne, 2004 for review; but see also Clegg & Legare, 2016). Additionally, children may deploy different social learning strategies depending on the complexity of the task. Thus, tasks that only assess learning by the presence or absence of target behaviors, rather than the ways in which children perform the actions, may not be sensitive to age- and task-related changes in social learning strategy use. Dickerson, Gerhardstein, Zack, and Barr (2013) developed a deferred imitation puzzle task that allows for a more sensitive measure of age- and task-related changes in memory and social learning strategy use. It has an *arbitrary sequence* structure, such that any order of actions will lead to successful reproduction of the overall goal. Arbitrary sequences are more difficult for toddlers to imitate than *enabling sequences* where only one order produces the desired outcome (Barr & Hayne, 1996; Bauer & Shore, 1987). It is not until 22 months of age that infants can imitate the order of an arbitrary sequence and 28 months of age when they can do so after a delay (Bauer, Hertsgaard, Dropik, & Daly, 1998). The arbitrary relational structure may impose constraints on recall by increasing the number of item relations that must be encoded, thus increasing memory load. In the puzzle task, memory load can be manipulated by increasing or reducing the number of to-be-remembered items or by altering visual complexity or perceptual cues of the task. This new puzzle task is less constrained compared to other imitation tasks because the magnetic puzzle pieces can be acted on in a variety of ways. The proportion of target behaviors relative to all behaviors produced during the puzzle task is measured to assess how faithful children are to the demonstrated gestures (action fidelity to index action emulation) and how efficiently they complete the puzzle (goal efficiency to index goal emulation).

2 | THE PRESENT STUDY

In the present study, to address whether memory load influences social learning strategies, groups of 2-year-old children were tested using established deferred imitation puzzle task protocols and were tested following a live demonstration (Dickerson et al., 2013; Zimmermann et al., 2015). Exp. 1 investigated the extent to which increasing the number of puzzle pieces would impact toddlers' imitation performance and the imitation strategies they used to complete the task. Exp. 2 examined imitation performance and strategies used when the perceptual features of the puzzle pieces differed. Reproduction of the demonstrated gestures (action emulation) and the final goal state (goal emulation) were coded in the present study. Performance was compared to baseline controls that did not observe a demonstration.

We included 2-year olds because children at this age are able to reproduce arbitrary sequences and imitate the puzzle task but performance is neither at ceiling nor floor (Dickerson et al., 2013; Zimmermann et al., 2015). Sex differences have been observed under some conditions with Dickerson et al.'s task (Zimmermann et al., 2015; Zimmermann, Moser, Gerhardstein, & Barr, in press), and so, we will test for sex differences in the present study.

Prior studies included two different puzzle stimuli, a boat and a fish. These studies (Dickerson et al., 2013; Zimmermann et al., 2015) found that the fish puzzle, because of its arbitrary and abstract shape boundaries, was more difficult than the boat. The boat was highly iconic and was divided along shape boundaries (i.e., the “sails” were triangles). Therefore, to maximize memory load, the present study deploys variants of the fish puzzle.

In addition to an arbitrary structure, the puzzle task also deliberately uses abstract objects, and its perceptual features can be adjusted to manipulate cognitive load. Item identity is more likely to be remembered than item order (Gulya et al., 1999), but as the number of items to remember increases, it becomes more difficult for infants to keep item identity in mind. In fact, Zosh and Feigenson (2012) found that 18-month olds were less likely to remember item identity as the number of items to be remembered (or memory load) increased. However, infants remembered more when the contrast between the features of the items increased. Thus, the addition of perceptual features may be helpful for imitation success and will be tested in Exp. 2.

3 | EXPERIMENT 1: MEMORY LOAD

Exp. 1 examined the role of memory load (as defined by number of puzzle pieces) on imitation of goals and gestures by 2-year olds using the puzzle imitation task introduced by Dickerson et al. (2013). To address whether memory load influences social learning strategies in the present experiment, toddler's imitation success with two-piece, three-piece, and three-piece plus distractor puzzles was tested. We wanted to examine whether there are changes in social learning strategy as a function of memory load. We predicted increased performance on the task with lower memory load. We had no specific prediction about how memory load would affect gesture performance, as the imitation literature has not often examined this aspect of behavior.

4 | METHODS

4.1 | Participants

The study included 66 typically developing children (38 boys) from two metropolitan areas. Children were tested at 2 years (*M* age = 24 months 18 days, *SD* = 16 days). Participants were primarily Caucasian (78.8%) and from college-educated families (*M* years of education = 17.22, *SD* = 1.45, as reported by 96.9% of families). The remaining 21.2% of the sample included the following races: Mixed (13.6%), African–American (4.5%), Asian (1.5%), and not reported (1.5%). Additionally, 9.1% of the sample was Latino. The mean rank of socioeconomic index (SEI; Nakao & Treas, 1994) was 75.94 (*SD* = 15.96) based on 46 families (70%). Following a partial replication design, an additional 15 children who participated in the three-piece plus distractor fish puzzle condition from Dickerson et al. (2013) were also included in the data analysis. Additional children were excluded from the analysis for the following reasons: five due to experimenter or technical error, five for failure to interact with the experimental stimuli, one due to parental interference, and seven for interacting with stimuli during the demonstration phase or failing to attend to the stimuli during the demonstration phase.

4.2 | Apparatus

The present study used a metal board inserted into a rectangular black case. The case was 35 cm tall, 42 cm wide, and 23 cm deep. The metal board could be easily slid in and out of the black case. The metal board was completely school bus yellow.

4.3 | Stimuli

The stimuli consisted of two, three, or four magnet pieces that were various shapes and colors but were the same thickness (.5 cm). These magnets were strong enough so that they stuck to the metal board, but they were weak enough so that they could be easily moved around. The pieces, when moved and connected correctly, formed a “fish.” At the beginning of the trial, each piece was placed in a different corner of the metal board. For each puzzle, there were two predetermined placements for the pieces. The puzzle pieces were all cut along a vertical orientation, and each separate piece was a different color (Figure 1).

4.4 | Design

Children were randomly assigned to independent groups in order to conduct a 3 (Memory load: 2-item, 3-item, 3-item + distractor) × 2 (Condition: experimental or baseline) between-subjects design. Within each condition, the arrangement of the puzzle pieces on the magnet board at the start of demonstration was counterbalanced across participants in one of two predetermined positions.

4.5 | Procedure

All protocols were approved by the Georgetown and Binghamton University IRBs. Testing primarily occurred in the home and a small subset (*n* = 19) was tested in the laboratory. The protocol was described to parents prior to obtaining informed consent from all parents. All of the children in the study were given a brief (5–10 min) warm up play session to ensure that they were familiar and comfortable with the experimenter. The apparatus was placed on a small table about one foot high. Before the task began, a black cloth covered the apparatus.

4.5.1 | Demonstration

The pieces were placed on the board behind the black cloth. The experimenter lifted the cloth and showed the toddler how to put the magnet pieces together to make the “fish.” The experimenter slid each piece by putting two fingers on the center of the piece. Every time a piece was moved, the experimenter made non-descriptive, fully scripted comments (“Look at this!,” “What was that?,” and “Isn't that fun?”) to orient the child to the demonstration. After moving the pieces into place to create the “fish,” the experimenter covered the apparatus with the black cloth and moved the pieces back into their original locations. The demonstration was conducted three times in total; the

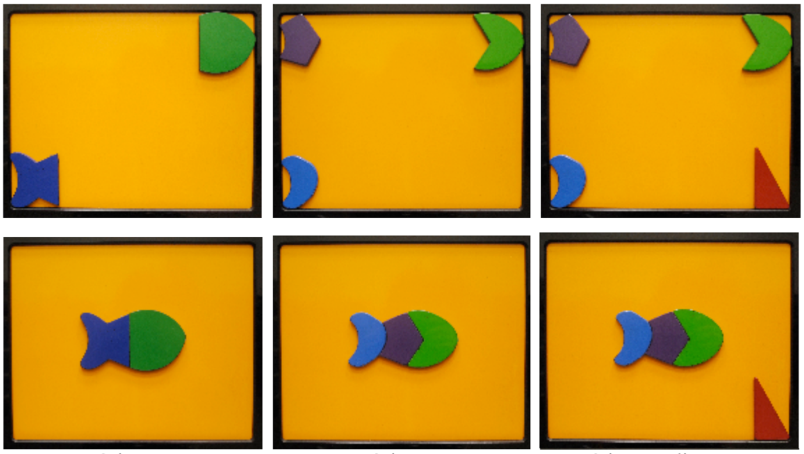

FIGURE 1 The top panel shows the disassembled puzzle pieces and the bottom panel shows the puzzles assembled. The puzzles as a function of increasing memory load. Left: low load 2-item puzzle. Middle: medium load 3-item puzzle. Right: high load 3-item plus distractor puzzle

three demonstrations together lasted approximately 50 s (M = 50.37 s, SD = 9.74 s). After the demonstration was finished, the experimenter covered the apparatus with the black cloth again and placed the pieces back into their original locations.

4.5.2 | Test phase

A short delay occurred between the end of the demonstration and the start of test to reset the pieces on the magnet (M = 6.69 s, SD = 2.17 s). The experimenter then lifted the black cloth up away from the apparatus and told the child "Now it's your turn!" The test lasted 60 s from the first time the child touched the magnet board or any of the magnet pieces (Figure 2). Following the 60 s test period, the experimenter conducted a manipulation check (demonstrated the target actions one time) and then gave the child the opportunity to reproduce them. The purpose of the manipulation check was to confirm that children were capable of sliding the puzzle pieces. As part of the manipulation check, experimenters asked the child, "What did you make?" to assess whether children could identify the final puzzle state as a fish. The purpose of the baseline was to assess whether children spontaneously produced the target gestures or goal of connecting the puzzle pieces when they were presented with the stimuli without a demonstration. Children in the baseline group did not receive a demonstration session and thus were not exposed to the stimuli until the start of the test phase.

FIGURE 2 Image of a 2-year-old child completing the fish puzzle in the high iconicity manipulation (Exp. 2)

4.6 | Coding

Imitation is operationally defined as duplicating the demonstrated actions at a rate significantly above baseline (Barr & Hayne, 2000). A detailed coding scheme was used to code the target behaviors (Dickerson et al., 2013; Moser et al., 2015; Zimmermann et al., 2015).

4.6.1 | On-task behaviors

Each contact with a puzzle piece (beginning when a piece was touched and ending when the touch ended) was coded. Each contact was coded along two dimensions: gesture and goal. On-task behaviors excluded exploratory play (interactions where the piece was removed from the board for more than 3 s) and micro-gestures (a piece was "nudged," meaning that it was moved less than 1/6 of the board) that did not result in any type of connection.

4.6.2 | Gesture coding

Coded actions included the following categories of gestures: correct slide, incorrect slide, strategy switch, and pick up and move.

4.6.3 | Goal coding

Coded actions that connected puzzle pieces included the following categories of goals: correct connection, target error connection, and connect other. Based on 30% of all test sessions rescored by a second coder, inter-rater reliability was very good ($\kappa_{gesture}$ = .77, κ_{goal} = .82).

The coded goals and gestures were used to compute four dependent measures (gesture imitation, action fidelity, goal imitation, and goal efficiency). Analyses of gestures to index action emulation via gesture imitation and action fidelity (both derived from gesture-coded actions) are presented first followed by analyses of goals to index goal emulation via goal imitation and goal efficiency (both derived from goal-directed actions). The coding of action fidelity and goal efficiency is included to more precisely characterize the participants' overall behavior during the test phase.

4.7 | Gesture imitation score

Following Dickerson et al. (2013), children received credit for each target puzzle piece that they correctly slid, up to a maximum of the number of target pieces for that puzzle, during the 60 s test period. The resulting gesture imitation score was then converted to a proportion to allow for cross measure comparison. No additional points were given for multiple correct slides with the same puzzle piece.

4.8 | Action fidelity score

To assess the rate at which correct slides were reproduced relative to other, less faithful actions, an action fidelity measure was calculated by taking the sum of all correct slides produced in the testing period (prior to reset following first puzzle completion) and dividing by all on-task behaviors produced during the 60 s test period (prior to reset following first puzzle completion). Higher proportions indicate more faithful reproduction of demonstrated actions; lower proportions indicate that increasing numbers of non-demonstrated actions were produced during the test (Moser et al., 2015; Zimmermann et al., 2015).

4.9 | Goal imitation score

Following Dickerson et al. (2013), children received one point for each correct connection. The maximum score for the 3-item and 3-item + distractor puzzles was 2; the maximum score for the 2-item puzzle was 1. As with the gesture imitation score, the goal imitation score was then converted to a proportion (either out of one or two). The goal imitation score is distinct from the gesture imitation and the action fidelity scores in that if a child used an incorrect gesture to correctly connect two puzzle pieces, they still received a point for the goal.

4.10 | Goal efficiency score

To assess the efficiency of piece connections relative to other on-task behaviors, a goal efficiency score was calculated as all correct connections performed as a proportion of all on-task behaviors across the 60 s test period or prior to first puzzle completion. Higher proportions indicate highly efficient puzzle reproduction; lower proportions indicate that increasing numbers of non-demonstrated actions were produced during the test (Moser et al., 2015; Zimmermann et al., 2015). For example, for the 2-item fish puzzle the child might simply move the front of the fish and connect it to the back without moving the back to most efficiently complete the puzzle, but another child might imitate by moving each piece toward the middle as demonstrated in order to connect them. Much less efficiently, another child may produce 20 on-task behaviors in the course of making the puzzle.

5 | RESULTS AND DISCUSSION

5.1 | Data analysis plan

Baseline was, as expected, zero for all participants for both gesture imitation and goal imitation scores. Due to the lack of variance in the baseline, non-parametric chi square (χ^2) or single sample t-test comparisons to zero are used to compare performance to baseline. We next assessed whether performance between groups differed as a function of memory load (2-, 3-, or 3-item + distractor puzzles) or sex of the child.

5.2 | Gesture imitation and action fidelity analysis

5.2.1 | Gesture imitation

A 3 (memory load: 2-item, 3-item, 3-item + distractor) × 2 (sex of child: boys, girls) ANOVA (analysis of variance) on gesture imitation yielded no main effect of memory load, $F(2,38) = 1.99$, $p = .15$, $\eta^2_p = .10$, or sex difference, $F < 1$, and no significant interaction, $F(2,38) = 2.84$, $p = .07$, $\eta^2_p = .13$. Follow-up non-parametric χ^2 were conducted to assess whether experimental groups differed from baseline. The 2-piece puzzle experimental group, $\chi^2(2) = 7.30$, $p = .03$, and the 3-item group, $\chi^2(2) = 12.54$, $p = .006$, were significantly different from baseline, but the 3-item + distractor was not, $\chi^2(2) = 3.80$, $p = .28$. See Table 1 for means.

5.2.2 | Action fidelity

A 3 (memory load: 2-item, 3-item, 3-item + distractor) × 2 (sex of child: boys, girls) ANOVA on action fidelity yielded no main effect of memory load, $F(2,38) = 1.00$, $p = .38$, $\eta^2_p = .50$, or sex difference, $F < 1$, and no significant interaction, $F(2,38) = 1.59$, $p = .22$, $\eta^2_p = .08$. Follow-up single sample t-tests ($p < .05$) comparing experimental group performance to zero showed that all experimental groups had a significantly higher action fidelity score than zero, 2-piece puzzle, $t(16) = 3.46$, $p = .003$, 3-piece puzzle, $t(13) = 3.11$, $p = .01$, and 3-piece + distractor, $t(12) = 2.50$, $p = .03$. See Table 1 for means.

5.3 | Goal imitation and goal efficiency analysis

5.3.1 | Goal imitation

A 3 (memory load: 2-item, 3-item, 3-item + distractor) × 2 (sex of child: boys, girls) ANOVA on goal imitation yielded a main effect of memory load, $F(2,38) = 17.45$, $p < .001$, $\eta^2_p = .48$, and a sex difference, $F(1,38) = 6.39$, $p = .02$, $\eta^2_p = .144$, but no interaction, $F < 1$. Girls performed better than boys (Figure 3). Post hoc SNK tests ($p < .05$) indicated that goal imitation of the 2-item puzzle ($M = .82$, $SD = .39$) was significantly

TABLE 1 Gesture score (SD) and action fidelity (SD) as a function memory load

	2-item	3-item	3-item + distractor
Gesture score	.41 (.40)	.45 (.38)	.21 (.35)
Action fidelity	.22 (.26)	.29 (.34)	.14 (.21)

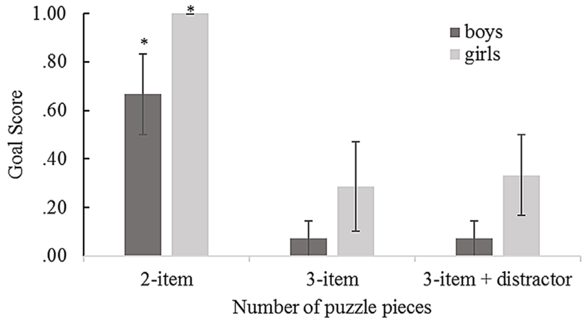

FIGURE 3 The goal imitation score (±1 SE) as a function of sex of the child and number of puzzle items. An asterisk indicates that the group goal score exceeded baseline

higher than the 3-item puzzle ($M = .18$, $SD = .37$) and the 3-item + distractor puzzle ($M = .19$, $SD = .35$), which did not differ from one another. Follow-up non-parametric χ^2 tests were conducted to assess whether groups differed from baseline. The 2-piece puzzle experimental group was significantly different from baseline, $\chi^2(1) = 15.90$, $p < .05$, but not the 3-item or the 3-item + distractor puzzle groups, $\chi^2(1) = 2.68$, $p = .26$ and $\chi^2(1) = 5.06$, $p = .08$, respectively (Figure 3).

5.3.2 | Goal efficiency

A 3 (memory load: 2-item, 3-item, 3-item + distractor) × 2 (sex of child: boys, girls) ANOVA on goal efficiency yielded a main effect of memory load, $F(2,38) = 8.10$, $p < .001$, $\eta^2_p = .30$, showing that as memory load increased, goal efficiency declined, and a sex difference, $F(1,38) = 9.43$, $p = .004$, $\eta^2_p = .20$; girls performed better than boys across all load levels. There was no significant interaction, $F(2,38) = 1.10$, non-significant (n.s.) (Table 2). Post hoc SNK tests ($p < .05$) indicated that goal efficiency of the 2-item puzzle ($M = .36$, $SD = .30$) was significantly greater than either the 3-item puzzle ($M = .09$, $SD = .21$) or the 3-item + distractor puzzle ($M = .09$, $SD = .26$), which did not differ from one another. Follow-up single sample t-tests ($p < .05$) comparing experimental group performance to zero showed that the 2-piece puzzle group performed significantly above zero, $t(16) = 4.81$, $p < .001$, but the 3-piece and 3-item + distractor did not, $t(13) = 1.72$, $p = .11$ and $t(12) = 1.97$, $p = .07$, respectively, Table 2.

6 | DISCUSSION

In Exp. 1, 2-year olds completed arbitrary puzzle sequences. Comparisons against baseline revealed that both gesture score and action fidelity were above baseline for the 2- and 3-item group, and that for the 3-item + distractor puzzle, action fidelity was above baseline. Children achieved higher goal scores when the sequence length was shorter. Furthermore, comparison to baseline revealed that the goal imitation and goal efficiency scores only exceeded baseline for the 2-item puzzle. These results where goal scores did not exceed baseline for higher sequence lengths are consistent with those of past studies using the puzzle task where performance on the fish task was also generally poor in 2-year olds (Dickerson et al., 2013; Zimmermann et al., 2015). Taken together these findings suggest that gesture reproduction precedes goal imitation. That is, copying how a task is performed may initially be easier to learn than the goal. This outcome is consistent with the hypothesis of Want and Harris (2002) that gesture imitation emerges before goal imitation. The effect of sequence length on goal performance is consistent with others' reports using enabling sequences in infants (Bauer & Fivush, 1992; Bauer & Mandler, 1992; Mandler & McDonough, 1995), and arbitrary sequences in 3-year olds (Subiaul & Schilder, 2014). Specifically, Subiaul and Schilder (2014) documented higher imitation sequence scores on 2-item than 3-item sequences on a touchscreen imitation task. In the present study, although the goal imitation score was high for the 2-item puzzle, overall goal efficiency and action fidelity were not. Unlike prior investigations, however, boys had significantly lower goal imitation and goal efficiency scores than girls, a difference that we will consider further in the general discussion. In Exp. 2, we explored perceptual factors that might influence cognitive load and subsequent imitation performance as indexed by our different measures.

TABLE 2 Goal efficiency (SD) as a function of sex of the child and number of puzzle pieces

	2-item	3-item	3-item plus distractor
Goal efficiency			
Boys	.19 (.17)	.04 (.09)	.02 (.05)
Girls	.52 (.33)	.15 (.27)	.17 (.21)

7 | EXPERIMENT 2: IMPACT OF PERCEPTUAL CUE CHANGES ON IMITATION

Factors that may influence memory load extend beyond item number. A large number of potential cue changes, relative to the puzzle configurations used in Exp. 1, are possible given the nature of the puzzle task. Exp. 2 tests two such changes: altering boundary cues and addition of surface feature cues (e.g., fish fins, eye, and ocean background context). The initial 2-item puzzle test in Exp. 1 used a vertically oriented straight cut, with colors that differed between pieces, potentially making the perception of the completed puzzle as a singular object more difficult. The presence of two or more colors (within a single piece), consistent or inconsistent with piece boundaries, may influence whether or not the puzzle is completed by young children. Thus, we manipulated perceptual features by changing the orientation of the puzzle piece cuts to a horizontal cut, so that color cues could be matched across pieces. To increase the potential impact of the cut manipulation, we created horizontal cut pieces that reflected *bilateral symmetry*, a feature that Levine, Huttenlocher, Taylor, and Langrock (1999), using a spatial transformation task with children 4- to 6-years old, found to facilitate performance.

Additionally, we tested the effect of adding additional surface feature cues to the previously tested puzzle pieces. Prior studies with arbitrary sequences have included additional semantic cues which have been shown to provide additional retrieval benefits (e.g., Zimmermann et al., 2015). Zimmermann et al. (2015) found that 2-year olds were more likely to make a 3-piece boat, but not a 3-piece fish puzzle when a semantically relevant context (the ocean) was added to the puzzle board. Research also suggests that iconicity facilitates learning in infancy, as 1.5–2-year olds are better at recognizing images rich in details relative to abstract shapes or objects lacking details (Pereira & Smith, 2009). Two-year olds also perform more target actions from a picture book when drawings are colorful and iconic relative to less iconic black and white line drawings (Simcock & DeLoache, 2006). We examine whether embedding the abstract puzzle pieces in a more meaningful context will enhance performance on the 2-item fish puzzle by adding internal feature cues to the puzzle pieces (e.g., fins, scales, and eye), against a semantically meaningful ocean background (e.g., water, waves, and sun).

7.1 | Participants

The study included 54 typically developing children (27 boys) from two metropolitan areas. Children were tested at 2 years (*M* age = 24 months 21 days, *SD* = 16 days). Participants were primarily Caucasian (74.0%) and from college-educated families (*M* years of education = 17.58, *SD* = 1.06, 98% reporting). The remaining sample included the following races: mixed (22%) and African–American (<1%). Additionally, 13% of the sample was Latino. The mean rank of socioeconomic index (SEI; Nakao & Treas, 1994) was 74.63 (*SD* = 13.78) based on 41 families (76%). Additional children were excluded from the analysis for the following reasons: two due to experimenter error, two for failure to interact with the experimental stimuli, three due to parental interference, six for interacting with the stimuli prior to test or failing to look for at least half of the demonstration, one for equipment issues, and two because the data could not be coded. Performance was compared to the vertical cut 2-item puzzle group from Exp. 1.

7.2 | Apparatus

The apparatus was the same as Exp. 1.

7.3 | Stimuli

The stimuli were constructed and start locations determined in the same manner as in Exp. 1. The 3-piece puzzle from Exp. 1 was cut horizontally to create the 2-item bilateral symmetry puzzle, which had three colors with a horizontal cut across color boundaries resulting in high levels of color boundary segmentation (Figure 4, left panel). The bilaterally symmetric version is more typical of commercial puzzles, which are typically cut across color boundaries. The 2-item iconicity puzzle (iconic) was created by adding internal features and a background to the vertical two-piece “fish” puzzle from Exp. 1. The semantically relevant background was added to the magnet board as in Zimmermann et al. (2015). In the iconic condition, the metal board displayed a cartoon of the ocean. The caricature of the ocean had a light blue sky, with dark blue waves representing the ocean, and a yellow sun located at the center left of the sky. The sun was composed of one semi-circle and three triangles (Figure 4, right panel). The data from Exp. 1 (from the vertical-cut 2-item puzzle, which had two colors with a vertical cut separating the colors between the head and tail), were used for comparison purposes (Figure 4, middle).

7.4 | Design

Children were randomly assigned to independent groups in order to conduct a 3 (Puzzle type: bilateral symmetry, vertical, iconic) × 2 (Condition: experimental or baseline) between-subjects design. Using a cross-experiment comparison, data from the 2-item vertical cut experimental and vertical cut baseline came from Exp. 1. Girls and boys were randomly assigned to conditions, to permit an explicit analysis of sex, following the finding of a sex difference in Exp. 1.

7.5 | Procedure

The same procedure as in Exp. 1 was followed including collection of the demographic measures. Experimental conditions included both the demonstration and test phases as before. The three demonstrations together lasted approximately 45 s (*M* = 43.70 s, *SD* = 6.99 s). There

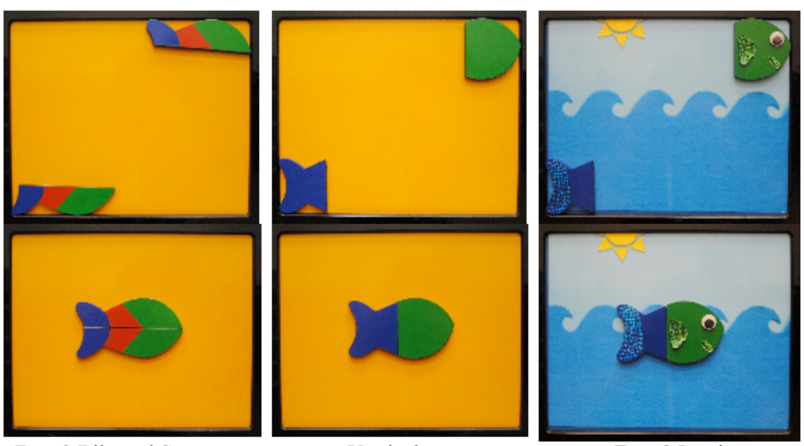

FIGURE 4 Image of the puzzle pieces used in Exp. 2 from left to right for the bilateral symmetry, vertical (from Exp. 1) and high iconicity (iconic) conditions

was a short delay between the end of the demonstration and the start of test to reset the pieces on the magnet (M = 7.00 s, SD = 5.48 s).

8 | RESULTS

8.1 | Data analysis plan

Baseline was again zero for all participants for both gesture imitation and goal imitation scores and therefore non-parametric χ^2 or single sample t-test comparisons to zero were used to compare performance to baseline. Next, we assessed whether performance between groups differed as a function of puzzle type (bilateral symmetry, vertical, or iconic) or sex of the child.

8.2 | Gesture imitation and action fidelity analysis

8.2.1 | Gesture imitation

A 3 (puzzle type: bilateral symmetry, vertical, iconic) × 2 (sex of child: boys, girls) ANOVA on gesture imitation yielded a main effect of puzzle type, $F(2,41) = 5.19$, $p < .01$, $\eta^2_p = .20$, no sex difference, $F(1,41) = 1.990$, $p = .16$, and no significant interaction, $F < 1$. Post hoc SNK tests comparing the different puzzle types revealed that children in the iconic puzzle condition performed as well as those in the vertical cut puzzle condition. Both of these groups produced significantly more gestures than those in the bilateral symmetry condition. Follow-up non-parametric χ^2 test indicated that the bilateral symmetry group did not differ from baseline, $\chi^2(1) = 2.84$, $p = .24$, however, the iconic group performed significantly above baseline, $\chi^2(1) = 6.14$, $p = .047$. Taken together, these results suggest that gesture performance was lower in the bilateral symmetry condition than in the other two conditions (Table 3).

8.2.2 | Action fidelity

A 3 (puzzle type: bilateral symmetry, vertical, iconic) × 2 (sex of child: boys, girls) ANOVA on action fidelity yielded a marginal main effect of puzzle type, $F(2,41) = 3.17$, $p = .052$, $\eta^2_p = .13$, but not sex, $F < 1$, and no significant interaction, $F < 1$. A follow-up SNK test, however, did not reveal differences in individual group comparisons. Single sample t-tests comparing experimental group performance to zero showed that the iconic group performed significantly above baseline, $t(14) = 2.84$, $p = .01$, but the bilateral symmetry puzzle group did not differ from zero, $t(14) = 1.39$, $p = .19$, in agreement with the outcome of the analysis of gesture imitation scores (Table 3).

8.3 | Goal imitation and goal efficiency analysis

8.3.1 | Goal imitation

A 3 (puzzle type: bilateral symmetry, vertical, iconic) × 2 (sex of child: boys, girls) ANOVA on goal imitation yielded no effect of puzzle type, $F < 1$, or sex, $F < 1$, and no interaction, $F < 1$. A follow-up non-parametric χ^2 test revealed that the bilateral symmetry puzzle differed from baseline, $\chi^2(1) = 11.14$, $p < .001$, as did the iconic puzzle group, $\chi^2(1) = 12.78$, $p < .001$ (Table 3).

8.3.2 | Goal efficiency

A 3 (puzzle type: bilateral symmetry, vertical, iconic) × 2 (sex of child: boys, girls) ANOVA on goal efficiency yielded no effect of puzzle type or sex, F's < 1. A significant interaction was present, $F(2, 41) = 3.54$, $p = .04$, $\eta^2_p = .14$. To assess the interaction effect, simple effects (follow-up) ANOVAs for each sex were conducted. There was no reliable effect for girls, $F(2,41) = 1.59$, $p = .21$. The effect for boys was close to significance, however, $F(2,41) = 2.59$, $p = .08$, suggesting that the interaction stems from a different pattern of results across the puzzle types in boys, but not in girls. Follow-up single sample t-tests ($p < .05$) comparing experimental group performance to zero showed that both the bilateral symmetry and iconic groups (not broken out by sex of child) had goal efficiency scores significantly higher than zero; $t(14) = 4.55$, $p < .001$ and $t(14) = 3.93$, $p < .005$, respectively (Table 3).

9 | DISCUSSION

We predicted, from the findings of Levine et al. (1999), that bilateral symmetry would assist in learning the actions associated with construction of a puzzle. We also predicted, from Periera and Smith (2009), that adding internal cues and a context (the iconic condition) would increase performance. These predictions were not borne out by goal scores, which were uniformly above baseline across sex and puzzle type, but the findings from other measures suggest that boys derived more assistance from the augmentations provided in the iconic and bilateral symmetry conditions than did girls. For boys, goal efficiency in the vertical-cut condition was lower than in both the bilateral symmetry and iconic conditions.

Additionally, gesture imitation and action fidelity did not exceed baseline levels for the bilateral symmetry group. There was no sex difference; both boys and girls showed a decline. It is possible that the decline occurred because moving the horizontally cut pieces (in the bilateral symmetry puzzle) was not as easy as moving the vertically cut pieces. While the current scoring system was not intended to code for this type of behavior, examination of the videos from the bilateral symmetry condition strongly supports this interpretation; many of the children in the bilateral condition grasped and removed pieces from the board, rather than imitating the demonstrated sliding gesture. Finally, the perceptual cues may have also resulted in changes in task performance in the bilateral symmetry group. Pereira and Smith (2009) reported that conflicting global and local segmentation cues were problematic for 2-year olds; if the finer details of the object (local cues) did not match with the overall shape of the object presented (global cue), children had a more difficult time in recognizing the objects. In the present experiment, our perceptual manipulation resulted in color boundaries that differed as a function of the cut as well. Additional research is needed to test differences in performance as a function of local and global cue perceptual changes.

Reduced gesture imitation and action fidelity by the bilateral symmetry group accompanied by generally high goal scores reflects a

TABLE 3 Means (SD) as a function of sex of the child and experimental condition in Exp. 2

N	Sex of child	Puzzle type	Gesture imitation	Action fidelity	Goal imitation	Goal efficiency
7	Boys	Bilateral symmetry	.07 (.19)	.02 (.06)	.71 (.49)	.55 (.46)
9		Vertical	.33 (.35)	.16 (.19)	.67 (.50)	.19 (.17)
8		Iconic	.25 (.27)	.21 (.36)	.75 (.46)	.46 (.40)
8	Girls	Bilateral symmetry	.06 (.18)	.04 (.12)	.75 (.46)	.38 (.34)
8		Vertical	.50 (.46)	.30 (.33)	1.00 (.00)	.52 (.33)
7		Iconic	.50 (.41)	.26 (.28)	.71 (.49)	.21 (.23)
12	Total	Bilateral symmetry	.07 (.18)	.03 (.09)	.73 (.46)	.46 (.39)
17		Vertical	.41 (.40)	.22 (.26)	.82 (.39)	.35 (.30)
15		Iconic	.37 (.35)	.23 (.32)	.73 (.46)	.34 (.34)

Note the data for the vertical group was collected in Exp. 1 and is analyzed as a cross-experimental comparison.

shift in social learning strategy to an emulation strategy. This strategy shift is consistent with Nielsen's (2006) prediction that strategy shift should vary as a function of task demands rather than changes in strategy use occurring purely as a function of age (see also Over & Carpenter, 2012). There might be a trade-off between instrumental learning and social strategy deployment depending on the affordances of the objects and in the current study toddlers were able to act appropriately based on the object affordances.

10 | GENERAL DISCUSSION

Children in Exp. 1 demonstrated a clear effect of cognitive load; 2-item puzzles were imitated relatively accurately, while 3-item puzzles both with and without a distractor resulted in much less accurate imitation. These results suggest that the constraints on 2-year olds' memory development, specifically working memory under conditions of abstract sequence memory load, may have limited their performance in Exp. 1. Further, testing in Exp. 1 showed that, at all levels, boys performed lower than girls in terms of puzzle construction (goal imitation and efficiency), but not in terms of gesture imitation. Altering the puzzle cut to include bilateral symmetry and boundary cues (Exp. 2) did not change performance (overall goal imitation) across conditions. In the bilateral symmetry condition, children noticed the affordances of the larger puzzle pieces and did not attempt to imitate the demonstrated gestures, yet completed the puzzle. However, different patterns of results were observed for boys and girls on goal efficiency. Boys showed a marginally significant performance increase in the bilateral symmetry condition, while girls did not. When internal surface features were added (Exp. 2, iconic condition), the same pattern was observed: girls' performance did not change significantly, while boys' goal efficiency performance marginally improved.

In addition to its use as a memory testing protocol, deferred imitation is used to assess social learning across early childhood. For example, imitation is the mechanism by which humans copy others to maintain a social interaction (Meltzoff, 2007; Nielsen, 2006; Nielsen & Tomaselli, 2010; Over & Carpenter, 2012; Uzgiris, 1981). Through imitation, children can learn both about instrumental or goal-directed actions, and conventional rituals including sometimes unusual or even unnecessary gestures (Clegg & Legare, 2016). Clegg and Legare argue that early in development children are very sensitive to the cues available in the environment and selectively imitate based upon those cues. Children may selectively omit parts of a sequence when the goal is clear. Under ambiguous learning conditions, precise imitation of each action may provide the child with pivotal information about how things work. Under these conditions, imitating the *how* may precede learning the *goal* of a sequence of actions; that is, in order to learn about their environment, children may first watch how others act on the environment and then do the same in order to achieve a goal. Our Exp. 1 findings that increases in gesture and action fidelity performance preceded goal imitation for the higher memory load puzzles are consistent with this argument. Furthermore, goal efficiency scores on the 2-piece puzzles suggest that, even in conditions where the children performed well (as measured by goal score), their actions are not those of an observer who directly acquired and imitated the demonstrated actions. Boys showed an increase in efficiency for the bilateral symmetry cut over the vertical cut puzzle; this increase may have resulted from a tendency (in boys) to simply move one piece into the other without moving the second, an action that will increase the goal efficiency score. Children in all groups and across both experiments were relatively inefficient in their imitation actions, even when their goal scores were quite high, suggesting that a trade-off may take place between instrumental and conventional imitation learning under ambiguous learning conditions (Clegg & Legare, 2016). Taken together, the findings demonstrate the role of rapid changes in the developing perceptual and memory systems on the deployment of social learning strategies.

Social learning, as assessed by the gesture imitation and action fidelity scores, reflected consistently poor performance. These metrics provided a different perspective, enabling examination of spatial skill, and what Want and Harris hypothesize as the different levels of social learning. Toddlers are making a large number of errors along the way to successful emulation (Want and Harris' term), and in many cases, these errors preclude actual "imitation," even though the children are able to accomplish the goal eventually and sometimes the demonstrated gestures as well.

The results for the perceptual changes made in Exp. 2 were mixed. Both types of perceptual cue change augmented efficiency of imitation by boys, but not girls. While we must interpret these findings cautiously due to the small sample of each sex by condition cell, the different pattern of results suggest that boys and girls may be solving the puzzles in different ways. These findings are further supported by emerging evidence from tests of transfer between 2D learning situations and a 3D test at 24 months also showing sex differences (Zimmermann et al., 2015, in press). The perceptual enhancements were more likely to facilitate puzzle performance by boys than by girls, suggesting that girls may possess a greater proficiency with puzzles at 2 years in general. Thus, the boys were in a better position to benefit from the additional cues, due to their initially low efficiency.

Is there really a difference between girls and boys in puzzle play at the age tested? Existing studies report differences in spatially relevant input to girls and boys at school and at home. At school, preschool teachers are reported to spend more time with boys than girls in traditional games that involve spatial transformation including block construction, sand play, and climbing areas, and with girls in the dramatic play area, which does not involve spatial transformation (Ebbeck, 1984). Levine, Ratliff, Huttenlocher, and Cannon (2012) (see also Levine et al., 1999) found sex differences in spatial skill in preschoolers as indexed by performance on a spatial transformation mental rotation task. They observed children at home every 4 months between 2 and 4 years of age and assessed the frequency and quality of puzzle play during these free play sessions. They reported that differences in spatial skills were related to the frequency and quality of puzzle play that children experienced between 2 and 4 years of age. More frequent puzzle play during toddlerhood was associated with higher performance on the mental rotation task for boys and girls, although overall boys' performance exceeded that of girls. Furthermore, they found that children's puzzle play quality, as indexed by more complex puzzles, more frequent interaction, and more spatial language, was significantly higher for boys than girls between 2 and 4 years of age. The quality of puzzle play was associated with better spatial skills for girls but not boys at 4.5 years of age (Levine et al., 2012). The authors argued that the association was found for girls and not boys because there was more variability in puzzle play for girls. The authors speculated that because girls depend more on verbal strategies during spatial tasks, verbal scaffolding during puzzle play might be particularly beneficial to girls' spatial development (Levine et al., 2012).

In the present study, girls outperformed boys in Exp. 1, and in Exp. 2, perceptual enhancements facilitated boys' but not girls' performance. We wondered if puzzle play in the home might differ for boys and girls. We collected information on time spent in puzzle play (minutes per day), number of strategies used during puzzle play (e.g., rotating pieces, finding edges, matching colors), and puzzle types, which were largely either simple, semantically themed peg puzzles or jigsaw puzzles with predominantly fewer than 10 pieces. For the subset of parents in the study that we asked about home puzzle play use, we found that there was a significant difference in the amount of puzzle play between boys and girls, $t(41.96) = 2.62$, $p = .01$, with girls spending on average 33.5 min per day with puzzles, and boys spending an average of 17.5 min per day. There were no differences on our other measures. Both boys and girls played with peg puzzles and simple jigsaw puzzles approximately 50% of the time.

It is important to note that the sample sizes are low when the sex of the child is considered as a factor. The small sample sizes mean that the differences between boys and girls must be interpreted cautiously. A future meta-analysis of data collected with the puzzle task will enable us to better understand the relationship between puzzle imitation, transfer of learning, socio-economic status (SES), sex of the child, child vocabulary, and household learning factors (puzzle time and puzzle type). A meta-analysis should elucidate factors that might be predictive of puzzle play early in development. Understanding which factors predict spatial learning during early childhood has important implications for both parents and educators (Levine et al., 2012), and for parents of girls in particular. Regardless of the outcome of this analysis, the finding of sex differences (overall better performance by girls under some conditions and more puzzle play at 2 years of age) indicates that parents should be encouraged to continue engagement in puzzle play with both boys and girls across early childhood.

Since Rovee-Collier (1996) issued her call *to shift the focus from the what to the why of infant learning and memory*, research in those fields has expanded dramatically. Factors that influence memory load (presence of distractors, number of items to-be-recalled) influence the type of social learning strategy used as well as overall success at completing the puzzle. Reducing memory load within the same age appears to lead to social learning strategy shifts, from imitation of actions to emulation of goals (Want & Harris, 2002). The present study demonstrates that learning involves a number of highly orchestrated and interconnected motor, perceptual, memory, and social systems.

ACKNOWLEDGMENTS

Thank you to the families and children who made this research possible. Thank you to Christine Cha and Amanda Grenell for help in data collection and collation. This research was supported by an NSF Grant to Rachel Barr and Peter Gerhardstein (BSC-1023772). Thank you also to Carolyn Rovee-Collier, who encouraged us to think, to question, to design, and test our ideas and to work collaboratively.

REFERENCES

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.

Barnat, S. B., Klein, P., & Meltzoff, A. N. (1996). Deferred imitation across changes in context and object: Memory and generalization in 14-month-old infants. *Infant Behavior and Development*, *19*, 241–251.

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development*, *19*, 159–170.

Barr, R., & Hayne, H. (1996). The effect of event structure on imitation in infancy: Practice makes perfect? *Infant Behavior & Development*, *19*, 253–257.

Barr, R., & Hayne, H., (2000). Age-related changes in imitation: Implications for memory development. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Progress in infancy research* (pp. 21–67). Hillsdale, NJ: Lawrence Erlbaum Associates, Vol. *1*.

Barr, R., Rovee-Collier, C., & Learmonth, A. E. (2011). Potentiation in young infants: The origin of the prior knowledge effect? *Memory and Cognition*, *39*, 625–636. DOI: 10.3758/s13421-010-0037-0

Bauer, P. A., & Fivush, R. (1992). Constructing event representations: A foundation of variation and enabling relations. *Cognitive Development*, *7*, 381–401.

Bauer, P. J., Hertsgaard, L. A., Dropik, P., & Daly, B. P. (1998). When even arbitrary order becomes important: developments in reliable temporal sequencing of arbitrarily ordered events. *Memory*, *6*, 165–198. DOI: 10.1080/741942074

Bauer, P. J., & Mandler, J. M. (1992). Putting the horse before the cart: The use of temporal order in recall of events by one-year-old children. *Developmental Psychology*, *28*, 441–452.

Bauer, P. J., & Shore, C. M. (1987). Making a memorable event: Effects of familiarity and organization on young children's recall of action sequences. *Cognitive Development*, *2*, 327–338.

Clegg, J. M., & Legare, C. H. (2016). Instrumental and conventional interpretations of behavior are associated with distinct outcomes in early childhood. *Child Development*, *87*, 527–542.

Cowan, N (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavior and Brain Sciences*, *24*, 87–114.

Dickerson, K., Gerhardstein, P., Zack, E., & Barr, R. (2013). Age-related changes in learning across early childhood: A new imitation task. *Developmental Psychobiology*, *55*, 719–732. DOI: 10.1002/dev.21068

Ebbeck, M (1984). Equity for boys and girls: Some important issues. *Early Child Development and Care*, *18*, 119–131.

Flynn, E., & Whiten, A. (2013). Dissecting children's observational learning of complex actions through selective video displays. *Journal of Experimental Child Psychology*, *116*, 247–263. DOI: 10.1016/j.jecp.2013.06.001

Gulya, M., Rovee-Collier, C., Galluccio, L., & Wilk, A. (1998). Memory processing of a serial list by very young infants. *Psychological Science*, *9*, 303–307.

Gulya, M., Sweeney, B., & Rovee-Collier, C. (1999). Infants' memory processing of a serial list: List length effects. *J Exp Child Psychol*, *73*, 72–91. DOI: 10.1006/jecp.1999.2494

Hartshorn, K., Rovee-Collier, C., Gerhardstein, P., Bhatt, R. S., Wondoloski, T. L., Klein, P., & Campos-de-Carvalho, M (1998). The ontogeny of learning and memory over the first year-and-a-half of life. *Developmental Psychobiology*, *32*, 69–89.

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review*, *24*, 33–73.

Hayne, H. (2006). *Age-related changes in infant memory retrieval: Implications for knowledge acquisition. Processes of brain and cognitive development. Attention and performance XXI, 209–231*. Oxford: Oxford University Press.

Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: Age-related changes in deferred imitation. *Behavioral Neuroscience*, *114*, 77–83. DOI: 10.1037//0735-7044.114.1.77

Hayne, H., Rovee-Collier, C., & Perris, E. E. (1987). Categorization and memory retrieval in 3-month-olds. *Child Development*, *58*, 750–767.

Huang, T. C., & Charman, C. (2005). Graduations of emulation learning in infant's imitation of actions on objects. *Journal of Experimental Child Psychology*, *92*, 276–302.

Jones, S. S. (2007). Imitation in infancy: The development of mimicry. *Psychological Science*, *18*, 593–599. DOI: 10.1111/j.1467-9280.2007.01945.x

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society of London B Biological Science*, *364*, 2325–2335. DOI: 10.1098/rstb.2009.0045

Kressley-Mba, R., Lurg, S., & Knopf, M. (2005). Testing for deferred imitation of 2-and 3-step action sequences with 6-month-olds. *Infant Behavior and Development*, *28*, 82–86.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, *35*, 940–949.

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology*, *48*, 530–542. DOI: 10.1037/a0025913

Mandler, J. M., & McDonough, L. (1995). Long-term recall of event sequences in infancy. *Journal of Experimental Child Psychology*, *59*, 457–474. DOI: 10.1006/jecp.1995.1021

Meltzoff, A. N. (2007). 'Like me': A foundation for social cognition. *Developmental Science*, *10*, 126–134.

Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: Observational causal learning in 24-month-old infants. *Developmental Psychology*, *48*, 1215–1228. DOI: 10.1037/a0027440

Merriman, J., Rovee-Collier, C., & Wilk, A. E. (1997). Exemplar spacing and infants' memory for category information. *Infant Behavior and Development*, *20*, 219–232.

Moser, A., Gerhardstein, P., Zimmermann, L., Grenell, A., & Barr, R. (2015). Transfer learning from touchscreens in toddlers. *Journal of Experimental Child Psychology*, *137*, 137–155. DOI: 10.1016/j.jecp.2015.04.002

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488. DOI: 10.1037/h0045106

Nakao, K., & Treas, J. (1994). Updating occupational prestige and socioeconomic scores: How the new measures measure up. *Sociological Methodology*, *24*, 1–72.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, *42*, 555–565.

Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science*, *21*, 729–736. DOI: 10.1177/0956797610368808

Ohr, P., Fagen, J., Rovee-Collier, C., Hayne, H., & Vander Linde, E. (1989). Amount of training and retention by infants. *Developmental Psychobiology*, *22*, 69–80.

Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, *126*, 182.

Pereira, A. F., & Smith, L. B. (2009). Developmental changes in visual object recognition between 18 and 24 months of age. *Developmental Science*, *12*, 67–80. DOI: 10.1111/j.1467–7687.2008.00747.x

Rovee, C. K., & Rovee, D. T. (1969). Conjugate reinforcement of infant exploratory behavior. *Journal of Experimental Child Psychology*, *8*, 33–39.

Rovee-Collier, C. (1996). Shifting the focus from what to why. *Infant Behavior & Development*, *19*, 385–400.

Rovee-Collier, C. (1997). Dissociations in infant memory: Rethinking the development of implicit and explicit memory. *Psychological Review*, *104*, 467–498.

Rovee-Collier, C., Hayne, H., & Colombo, M. (2001). *The development of implicit and explicit memory*. Amsterdam: John Benjamins Publishing Co.

Shiffrin, R. M. (1993). Short-term memory: A brief commentary. *Memory and Cognition*, *21*, 193–197.

Simcock, G., & DeLoache, J. (2006). Get the picture? The effects of iconicity on toddlers' reenactment from picture books. *Developmental Psychology*, *42*, 1352–1357. DOI: 10.1037/0012-1649.42.6.1352

Subiaul, F., & Schilder, B. (2014). Working memory constraints on imitation and emulation. *Journal of Experimental Child Psychology*, *128*, 190–200.

Tennie, C., Call, J., & Tomasello, M. (2006). Push or pull: Imitation vs. emulation in great apes and human children. *Ethology*, *112*, 1159–1169.

Tennie, C., Greve, K., Gretscher, H., & Call, J. (2010). Two-year-old children copy more reliably and more often than nonhuman great apes in multiple observational learning tasks. *Primates*, *51*, 337–351. DOI: 10.1007/s10329-010-0208-4

Uzgiris, I. C. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development*, *4*, 1–12.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*, 163–167. DOI: 10.1016/j.cognition.2008.07.013

Want, S. C., & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. *Developmental Science*, *5*, 1–14. DOI: 10.1111/1467-7687.00194

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society of London B Biological Science*, *364*, 2417–2428. DOI: 10.1098/rstb.2009.0069

Wiebe, S. A., & Bauer, P. J. (2005). Interference from additional props in an elicited imitation task: When in sight, firmly in mind. *Journal of Cognition and Development*, *6*, 325–363.

Zimmermann, L., Gerhardstein, P., Moser, A., Grenell, A., Dickerson, K., Yao, C., & Barr, R. (2015). Do semantic contextual cues facilitate transfer learning from video in toddlers? *Frontiers in Psychology*, *6*, 561. DOI: 10.3389/fpsyg.2015.00561

Zimmermann, L., Moser, A., Gerhardstein, P., & Barr, R. The ghost in the touchscreen: Social scaffolds promote learning by toddlers. *Child Development* (in press).

Zosh, J. M., & Feigenson, L. (2012). Memory load affects object individuation in 18-month-old infants. *Journal of Experimental Child Psychology*, *113*, 322–336. DOI: 10.1016/j.jecp.2012.07.005