Journal of Experimental Child Psychology 214 (2022) 105286

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

# Two-year-olds use past memories to accomplish novel goals

Tashauna L. Blankenship [a,*], Melissa M. Kibbe [b,c]

[a] Department of Psychology, University of Massachusetts Boston, Boston, MA 02125, USA
[b] Department of Psychological & Brain Sciences, Boston University, Boston, MA 02215, USA
[c] Center for Systems Neuroscience, Boston University, Boston, MA 02215, USA

A R T I C L E   I N F O

*Article history:*
Received 8 January 2021
Revised 2 August 2021

*Keywords:*
Planning
Episodic memory
Generalization
Learning
Development
Episodic prospection

A B S T R A C T

Memory-guided planning involves retrieving relevant memories and applying that information in service of a goal. Previous studies have shown substantial development in this ability from 3 to 4 years of age. We investigated the emergence of memory-guided planning by asking whether 2-year-olds could draw on episodic memories of past experiences to generate and execute plans. In Experiments 1 and 2 ($N$ = 32, $ds$ > .7), 2-year-olds successfully did so, and this ability developed significantly across the third year of life. Furthermore, in Experiment 3 ($N$ = 19, $d$ = 0.63), 2-year-olds successfully applied episodic memories to guide plans in a novel problem context, suggesting flexibility in this ability. Together, these results suggest that some form of memory-guided planning emerges during the third year of life and may form the cognitive basis for episodic prospection later in development.



## Introduction

The ability to make a plan to accomplish a current or future goal is critical for typical functioning (Gollwitzer & Oettingen, 2011). Humans have the ability to consider hypothetical situations to make plans (Schacter & Addis, 2007; Suddendorf & Corballis, 2007), and doing so effectively is associated with academic success (Mau, 1995; Prabhakar, Coughlin, & Ghetti, 2016; Zhu & Mok, 2012). To make

* Corresponding author.
*E-mail address:* tashauna.blankenship@umb.edu (T.L. Blankenship).

https://doi.org/10.1016/j.jecp.2021.105286
0022-0965/

an effective plan, the planner must recall previous relevant knowledge and experiences and then apply these memories to a new as-yet-unexperienced situation. For example, when planning for their upcoming birthday party, children may recruit their existing knowledge of the structure of birthday parties (e.g., birthday parties typically involve a cake and presents) as well as remember past birthday parties they have attended (e.g., Julie had her birthday party at a pizza shop) in order to generate future plans (e.g., I will have cake, presents, and pizza at my birthday party next week). Thus, the planner can draw on both semantic (knowledge-based) and episodic (event-based) memory (Schacter & Addis, 2007; Schacter, Benoit, & Szpunar, 2017; Suddendorf & Corballis, 2007) when approaching a novel problem. We refer to the ability to draw on past knowledge and experiences to generate plans as memory-guided planning (Blankenship & Kibbe, 2019).

Memory-guided planning using episodic memory specifically is thought to be particularly crucial for typical cognitive functioning (Prabhakar et al., 2016) because it involves selecting relevant episode-bound experiences and applying those experiences to a new problem (Atance and O'Neill, 2001; Blankenship & Kibbe, 2019). Much of the work examining the development of memory-guided planning has focused on the development of episodic prospection in which children are asked to use episodic memories to come up with hypothetical future scenarios, an ability that is critical in adult planning (e.g., Schacter et al., 2017). In these studies, children are asked to imagine future scenarios at various time points and to provide a brief narrative of the events that they think will be likely to unfold, requiring children to retrieve episodic memories and apply those memories to an as-yet-unexperienced event (Busby & Suddendorf, 2005; Hayne, Gross, McNamee, Fitzgibbon, & Tustin, 2011). These studies have found that 3-year-olds have more difficulty than 4-year-olds in constructing narratives around future events, with children reliably using episodic memories to make future predictions by around 5 years of age (e.g., Atance & Meltzoff, 2005; Atance & O'Neill, 2005; Cuevas, Rajan, Morasch, & Bell, 2015; Suddendorf, Nielson, & Von Gehlen, 2011; Suddendorf & Redshaw, 2013). Indeed, episodic memory itself undergoes substantial development from 3 to 4 years of age (Hayne & Imuta, 2011; Scarf, Boden, Labuschagne, Gross & Hayne, 2017), and younger children also have fewer experiences and less preexisting knowledge (e.g., about trips to the beach) to draw on.

However, younger children have more success when they are asked to apply past experiences to *execute,* rather than *articulate,* a plan (Atance, Louw, & Clayton 2015; Blankenship & Kibbe, 2019; Prabhakar & Hudson, 2014; Suddendorf et al., 2011). For example, we previously examined 3- and 4-year-olds' capacity to deploy episodic memories to achieve novel nested goals (Blankenship & Kibbe, 2019). Children first interacted with a box that had four drawers embedded in it and were shown that the drawers could be opened by taking specific actions on four unique effectors on the box. During test trials, the drawers were baited with one to four colored beads, and children were asked to retrieve the beads in a specific order. Therefore, the task required children to retrieve the relevant episodic memories (how to open the baited drawers) and to make a plan to apply those episodes in the correct order. The 3-year-olds were able to successfully retrieve up to two beads in the correct order, and the 4-year-olds could successfully complete goals involving three or four beads. In another study, Suddendorf et al. (2011) examined children's ability to *generalize* past experiences to an entirely novel context, flexibly applying episodic memories in service of a related but superficially different problem (one of the requirements for episodic prospection). They showed 36- and 48-month-old children a puzzle box that could be activated by the insertion of a triangle-shaped key into its front. Children were then shown a new box with a square opening. Afterward, children were led to a different room and were given the opportunity to select a key to activate the new box from a set of different-shaped keys. Both 36- and 48-month-olds were able to do so if queried immediately following the box demonstrations, but only 48-month-olds were able to do so following a delay.

Together, these studies suggest that children as young as 3 years are capable of memory-guided planning when the cognitive demands of the task are reduced (e.g., when children are asked to take actions instead of articulate a plan, when memory demands are reduced by testing immediately following episodic memory formation). Previous work suggests that memory-guided planning undergoes significant development from 3 to 6 years of age when children begin to develop more adult-like episodic prospection (Atance & Meltzoff, 2005; Blankenship & Kibbe, 2019; Suddendorf et al., 2011). However, because much work has focused on children aged 3 years and over, the emergence of memory-guided planning is unknown. Previous work suggests that the cognitive components required to engage in memory-guided planning may be present by at least the second year of life, including explicit memory (Bauer & Dow, 1994; Bauer, Wiebe, Carver, Waters, & Nelson, 2003; Rovee-Collier & Sullivan, 1980), working memory (for a review, see Kibbe, 2015), and planning (e.g., McCarty, Clifton, & Collard, 1999; Stahl & Feigenson, 2015; Xu, Cote, & Baker, 2005). Children as young as 2 years (>26 months) are able to form explicit memories that include contextual information (e.g., location), suggesting that children younger than 3 years may be able to form episodic memories (Newcombe, Balcomb, Ferrara, Hansen, & Koski, 2014). Evidence from rational imitation tasks suggests that infants may be capable of action planning; after observing an adult taking an inefficient or irrelevant action on an object or tool, infants are more likely to imitate the *goal* of the action than the action itself (e.g., Gergely, Bekkering, & Király, 2002; Liszkai-Peres, Kampis, & Király, 2020; Schwier, van Maanen, Carpenter, & Tomasello, 2006), suggesting that infants can extract the goal structure of an event and use an unobserved (but efficient) means to accomplish that goal.

Given that many of the cognitive components required for memory-guided planning appear to be operational prior to 3 years of age, it is possible that the ability to coordinate these processes to generate effective plans may also be operational before age 3. We examined the emergence of the ability to draw on episodic memories to accomplish novel goals prior to age 3. In the current series of experiments, we asked whether 24- to 36-month-old toddlers could retrieve relevant episodic memories and apply them in service of goals both within the same context that the memories were acquired (Experiments 1 and 2) and within a novel context (Experiment 3). We used variations on our decision-based task previously used to examine memory-guided planning in 3- and 4-year-old children (Blankenship & Kibbe, 2019). Crucially, the tasks were designed to be appropriate for 2-year-olds; the tasks encouraged memory-guided planning while minimizing response demands and did not require children to have preexisting knowledge acquired outside of the laboratory.

In Experiment 1, we asked whether 2-year-olds could engage in memory-guided planning by drawing on episodic memories. We showed children a box containing two transparent drawers and demonstrated that each drawer could be opened by taking a unique action on the box (lifting a purple pom-pom or a pink cylinder). Children were then introduced to two animal characters, were told each animal's favorite color (red or yellow), and were shown that the animals' favorite colored beads could be retrieved from the drawers in the box. On each test trial, the drawers were baited with one yellow bead and one red bead, and children were asked to retrieve the favorite colored bead of one of the animals. Thus, the task required children to apply contextually based memories (how to open each of the drawers; requiring retrieval of episodic memory) and to select *only* the episode relevant to accomplish the novel goal (retrieving the correct bead for the animal character and not the other bead; requiring planning), engaging memory-guided planning. Critically, in Experiment 1 both drawers were always baited, but the color bead in each baited drawer varied across trials, as did the goal (i.e., which bead children were asked to retrieve). This meant that for children to respond correctly, they needed to retrieve the relevant memories (action that opens each drawer) to generate unique plans across trials. In Experiment 2, we aimed to conceptually replicate the results of Experiment 1 while reducing inhibitory control demands and controlling for a potential spatial confound.

In Experiment 3, we asked whether 2-year-olds could *generalize* from past episodic memories to an entirely novel context, demonstrating a more flexible use of memory-guided planning. We first showed children a blue-outlined box apparatus with a transparent drawer in the front. We then showed children that placing a yellow cylinder on top of the box made a purple bead appear in the drawer and placing a wooden block on top of the box made a green bead appear. We then switched the box and objects to similarly shaped, but perceptually distinct, box and objects—a yellow-outlined box, a pink cylinder, and a blue block—and asked children to retrieve specific beads from the box. We reasoned that if children are able to generalize past episodes to the new box and objects, then they should select the relevant shaped object when asked to retrieve a bead of a particular color despite never having interacted with the apparatus before.

## Experiment 1

### Method

#### Participants

Participants were 16 2-year-old children ($M$ = 30.1 months, range = 24.6–35.2; 6 girls). Sample size was determined prior to the experiment. A power analysis using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) suggested a total sample size of $N$ = 16 to observe a difference from chance using a one-sample $t$ test, assuming a large effect size ($d$ = 0.75, $\alpha$ = .05, $1 - \beta$ = .80). Parents identified their children as Black ($n$ = 1), Asian ($n$ = 3), White ($n$ = 8, of whom 2 were identified as Hispanic or Latinx), or multiracial ($n$ = 4). Furthermore, all parents self-reported that they had completed a college degree or higher. An additional 4 children were excluded because of experimenter error ($n$ = 3) or refusal to participate in the task ($n$ = 1). Children were recruited from the greater Boston area in the northeastern United States and received a small gift for their participation.

#### Materials

Materials included one 1-cm natural wood bead, 20 2-cm plastic beads (10 red and 10 yellow), two 15-cm stuffed animals (a monkey and a lion), and two 8-ounce clear plastic cups. Children interacted with a black rectangular box made from foam core board (53 × 17 × 16 cm). The box had two 5 × 5-cm openings in its front, bordered with colored tape (blue on left and green on right). Each opening contained a transparent plastic drawer that could be surreptitiously baited with beads by the experimenter through an opening in the back of the box. The experimenter opened each drawer by pushing the drawer out from the back, giving the appearance that the drawer opened in response to an action taken by the experimenter or the child. There were two objects affixed to the top of the box: a purple pom-pom attached to the box with a magnet on the left and two nested pink cylinders on the right. The pom-pom could be lifted off of its magnetic base. The inner cylinder could be lifted out of the outer cylinder. Children learned that each of these actions would open one of the transparent drawers. See Fig. 1 (left panel).

#### Procedure

Children were seated at a small table across from the experimenter with their caregiver seated in a chair behind them. The study session was recorded and stored for coding purposes. The task was separated into two phases: *encoding* and *test* (see Fig. 1, left panel).

#### Encoding phase

The experimenter first showed children the black box and told them that she was going to show them how it worked. The experimenter placed the plain wooden bead inside one of the drawers. She then said, “See this bead in here [pointing to the drawer]? Watch how I get it out.” She then demonstrated that performing an action on one of the objects on top of the box opened the drawer. If the bead was placed in the left drawer, the experimenter lifted and replaced the purple pom-pom from its magnetic base. If the bead was placed in the right drawer, the experimenter slid the inner pink cylinder in and out of the outer pink cylinder. The experimenter then prompted children to retrieve a bead from the drawer themselves. She again baited the drawer with the wooden bead and said, “Look! There’s a bead in here. I want to get it out. What should I do?” Children responded either by pointing to the relevant object ($n$ = 4) or by performing the action themselves ($n$ = 12). If children responded incorrectly, the experimenter showed that the action they selected did not work, re-demonstrated the correct object–action–drawer mapping, and then baited the drawer again and asked children to select the correct action to retrieve it.

Once children responded correctly, the experimenter demonstrated how to open the second drawer, first showing children the relevant object–action–drawer mapping and then baiting the drawer again and inviting them to point or take the action themselves. The order in which the object–action–drawer mappings were demonstrated was counterbalanced across participants; half the children saw the pom-pom lifting–left drawer mapping first, and the other half saw the cylinder

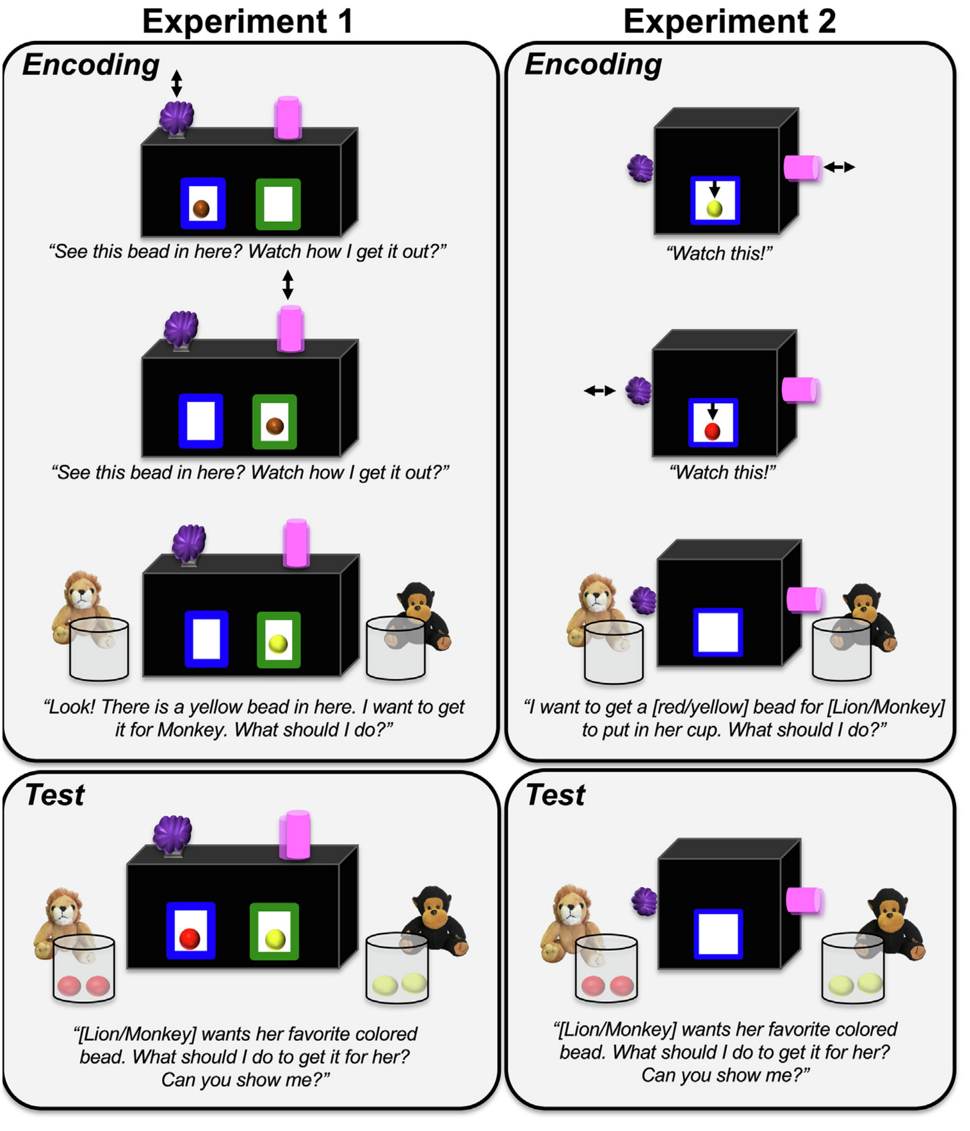



**Fig. 1.** Schematic representations of the encoding and test phases from Experiment 1 (left panel) and Experiment 2 (right panel). The arrows represent the actions taken on the objects (lifting and removing the pom-pom and sliding the inner pink cylinder in and out of the outer pink cylinder).

sliding–right drawer mapping first. We capped the number of times the experimenter re-demonstrated each object–action–drawer mapping at three per drawer. Thus, children could observe a maximum of four demonstrations of each object–action–drawer mapping. All children were able to respond correctly after fewer than the maximum of four demonstrations (*M* = 0.78 demonstrations, *SD* = 1.16).

Next, the experimenter introduced children to the animal characters. The experimenter first showed children one of the animals (which animal was presented first was counterbalanced across participants), placing the animal on one side of the box. She told children, "This is [Monkey/Lion]. [Monkey/Lion] loves the color [yellow/red]. [Yellow/Red] is her favorite color." The experimenter then placed a single red or yellow bead in one of the drawers and said, "Look! There is a [yellow/red] bead in here [*pointing to one of the drawers*]. I want to get it for [Monkey/Lion] to put in her cup. Do you think you can help me? What should I do?" If children responded correctly, the experimenter opened the drawer, retrieved the bead, and placed it in the transparent cup in front of the stuffed animal. If children selected the wrong action, children were shown the correct action and given an additional attempt (capped at four demonstrations per drawer). Once children successfully retrieved the bead, the experimenter baited the other drawer with the same color bead and asked children to retrieve the bead. The experimenter then introduced the second animal in the same manner, placing the animal on the right side of the box. Children were asked to retrieve two beads (one from each drawer) for the second animal. Children gave few incorrect responses with few repeat demonstrations needed ($M$ = 0.91, $SD$ = 1.38). All the beads that were retrieved during this phase were placed in transparent cups in front of each animal character. These visual reminders of each animal's favorite color were available during later test trials.

On completion of the encoding phase, all children had successfully retrieved three beads from each drawer themselves and had been familiarized with each animal's preferred color. The test phase began immediately following the encoding phase.

*Test phase*

During the test phase, children were again asked to retrieve a bead for one of the characters. However, during test trials, both drawers were baited—one with a yellow bead and one with a red bead. The experimenter then asked children to retrieve one of the beads, saying, "[Monkey/Lion] wants her favorite colored bead. What should I do to get it for her? Can you show me?" Critically, unlike during the encoding phase, the experimenter never directly indicated which drawer children should open; she did not refer to the color of the bead children were being asked to retrieve, nor did she point to a drawer. Because both drawers were baited, children could not simply select the action associated with one of the drawers. To select the correct action to open the drawer, children needed to recall the relevant episodes from the encoding phase (requiring memory retrieval) and then apply only the relevant action that would allow them to complete the goal of retrieving the correct bead (requiring planning).

Children received no explicit feedback from the experimenter during the test phase. If children selected a correct action, the experimenter opened the drawer, retrieved the bead, placed it in the animal's cup, said "Okay, let's try another one!" and reset the apparatus by removing the other bead through the opening in the back of the box. If children selected an incorrect action on a trial, the experimenter said "OK, let's try another one!" and reset the apparatus by removing both beads from the drawers through the opening in the back of the box (a bead was not placed in an animal's cup when children responded incorrectly).

Children completed two blocks of 4 trials each, with order of the trials counterbalanced in each block (e.g., Monkey, Lion, Lion, Monkey; Lion, Monkey, Monkey, Lion), for a total of 8 test trials. Whether the target bead was placed in the left or right drawer was counterbalanced across trials. Children were counted as responding correctly if they selected only the action that would open the correct drawer. If children selected the incorrect drawer, or if they switched their response, they were scored as choosing incorrectly.

*Results*

We first asked whether children selected the correct action to retrieve the target bead at rates above chance. We computed children's mean proportion correct across trials and compared these with chance (.50) using a one-sample $t$ test. We also used Bayes factor analysis to determine the likelihood of the alternative hypothesis (that children selected the correct action at rates above chance) versus the null hypothesis (that children selected the correct action at rates not different from chance) (see Gallistel, 2009; Jeffreys, 1961; Rouder, Speckman, Sun, Morey, & Iverson, 2009). We found that children selected the correct action at rates greater than would be expect by chance ($M$ = .63, $SD$ = .16), $t(15)$ = 3.04, $p$ = .01, $d$ = 0.76 (Fig. 2). Bayes factor analysis showed that the alternative hypothesis was nearly six times more likely than the null hypothesis ($BF_{10}$ = 5.88).

Even though children were not given explicit feedback, children who selected an incorrect action were able to observe that the action did not work and therefore may have formed and used low-level associations gleaned across the test trials, rather than memory-guided planning, to achieve above-chance performance. To examine this possibility, we asked whether children's performance increased across test trials or whether children's performance was consistent throughout. We conducted a binomial logistic regression on children's responses with trial (1–8) as a factor and participant ID as a covariate. If children's above-chance performance was driven by repeated exposure to the test trials, we would expect their performance to improve significantly across test trials. Contrary to this, we found no effect of trial ($\beta = -.01$, $p$ = .88). Children's mean performance on the first 4 trials also was not significantly different from their mean performance on the last 4 trials, $t(15) = -0.78$, $p$ = .45.

During encoding trials, children could observe anywhere from one to four demonstrations per object–action–drawer mapping. Therefore, we asked whether the number of demonstrations they observed during encoding was related to their performance on the test trials. We observed no correlation between number of demonstrations during the encoding phase and children's test performance ($r = -.03$, $p$ = .91).

Because the location of the beads was counterbalanced across trials, on half the trials the bead that children were being asked to retrieve was located on the same side as the relevant animal. For example, when children were asked to retrieve a bead for Monkey, on half the trials the bead was located on the same side of the box as Monkey and the correct action to open the drawer was located on the same side of the box as Monkey. On the other half, the bead was located in the drawer on the other side of the box. We asked whether children may have performed better when the correct action was located on the side adjacent to the relevant animal. A paired-samples $t$ test revealed that children responded similarly when the correct bead/action was located adjacent to the relevant animal ($M$ = .56, $SD$ = .21) compared with when the correct bead/action was located opposite the relevant animal ($M$ = .69, $SD$ = .23), $t(15) = -1.65$, $p$ = .12.

Finally, in an exploratory analysis, we asked whether children who chose correctly on the first test trial ultimately performed better overall than children who were not successful on the first trial. We ran an independent-samples $t$ test on children's mean proportion correct, comparing children who succeeded on the first test trial with those who did not. Results suggested that children who succeeded on the first test trial ($n$ = 11; $M$ = .66, $SD$ = .15) did not perform significantly better overall than those who did not ($n$ = 5; $M$ = .55, $SD$ = .19), $t(14)$ = 1.25, $p$ = .23, $d$ = 0.64. Similarly, performance on the first trial did not predict performance on the other 7 test trials, $F(1, 14)$ = 0.033, $p$ = .86, $R^2$ = .002. However, it is important to note that this analysis is exploratory and therefore should be interpreted with caution.

*Discussion*

In Experiment 1, we asked whether 2-year-old children could use episodic knowledge to plan for and accomplish goals, engaging in memory-guided planning. Children in Experiment 1 were able to successfully deploy the correct action to retrieve the target bead at rates above chance, suggesting that they recalled the relevant episode and applied that knowledge correctly in service of a goal. Because both drawers were baited with beads, and because we did not explicitly tell children which bead they were being asked to retrieve (i.e., we did not refer to the bead by its color but instead asked children to retrieve one of the animals' favorite colors), to succeed children needed to identify which drawer they needed to open to accomplish the goal (e.g., Lion likes red, so I need to open the drawer with the red bead) and to then select and apply only the relevant episode (e.g., lifting the pom-pom opens the drawer that currently contains the red bead) to develop an appropriate plan (e.g., engage in the action to open the drawer with the red bead but not the yellow bead) in order to achieve that goal (e.g., retrieve the red bead for Lion). Performance on the task did not improve across trials, suggesting that

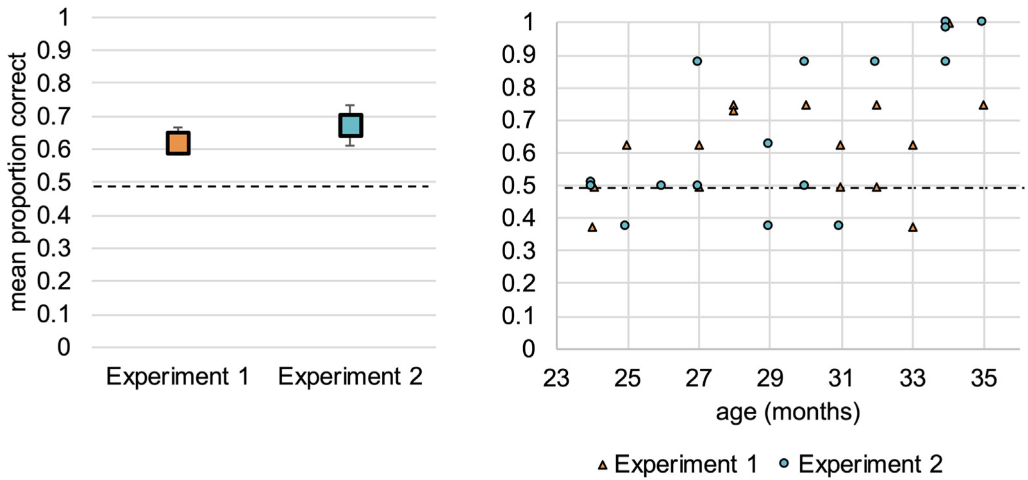



**Fig. 2.** Left panel: Overall mean proportions of correct responses for Experiments 1 and 2. Right panel: Individual children's mean proportions correct as a function of age for Experiments 1 and 2. Dashed lines show chance levels (.50).

children's ability to select the correct action did not improve with repeated exposure. These results suggest that the ability to draw on memories to guide plans may emerge prior to 3 years of age.

In Experiment 1, both beads were visible in the drawers throughout the task, which may have placed additional demands on children's executive functions; children not only needed to recall the relevant episode required to retrieve the correct bead on the trial but also needed to inhibit the irrelevant bead and its associated actions. Furthermore, the action objects were located in positions that were spatially adjacent to the drawers that they opened, which could have served as a lower-level cue to support children's ability to select the correct action. In Experiment 2, we aimed to conceptually replicate the results of Experiment 1 while reducing inhibitory control demands and eliminating these spatial cues. Experiment 2 used a similar but altered box apparatus compared with the one used in Experiment 1. In Experiment 2, children interacted with a box that had a single drawer in the front and the two action objects (the pom-pom and the cylinder) on either side of the box. Children learned that each action caused the box to release a different colored bead in the drawer. Thus, the drawer was empty until an action was taken on one of the objects, and there were no spatial associations between drawer and action objects. During the test phase, we again asked children to retrieve animals' favorite color beads. To succeed, children needed to remember the relevant action–bead pairing, and then deploy that knowledge to take the correct action.

## Experiment 2

### *Method*

#### *Participants*

A separate sample of 16 2-year-old children ($M$ = 29.9 months, range = 24.4–35.4; 6 girls) participated. Sample size was determined before the experiment using the same criteria as in Experiment 1. Parents identified their children as Black ($n$ = 1), White ($n$ = 10, of whom 2 were identified as Hispanic or Latinx), or multiracial ($n$ = 3) (2 parents did not report their children's race). Most parents self-reported that they had completed a college degree or higher. An additional 10 children were excluded because of experimenter error ($n$ = 4), recording equipment failure ($n$ = 1), refusal to participate in the task ($n$ = 2), exceeding four demonstrations during the encoding phase ($n$ = 2), or parental interference ($n$ = 1). Children were recruited from the greater Boston area and received a small gift for their participation.

#### *Materials*

Materials were similar to those in Experiment 1 except that the box apparatus had a single green-bordered 5 $\times$ 5-cm transparent drawer embedded in its front, and the action objects were located on the left and right sides of the box. See Fig. 1 (right panel) for a schematic representation of the apparatus.

#### *Procedure*

#### *Encoding phase*

The experimenter showed children the black box and told them that she was going to show them how it worked. The experimenter then demonstrated how taking actions on each of the objects on the box caused the box to release a different colored bead into the drawer (with order of demonstration counterbalanced across children). To demonstrate how to get the box to release a yellow bead, the experimenter lifted and removed the purple pom-pom, released a yellow bead into the drawer, and said "See [*holding up bead*]! This [*pointing to action object*] made a yellow bead come out!" To demonstrate how to make the box produce a red bead, the experimenter removed and reinserted the pink cylinder, released a red bead into the drawer, and said "See [*holding up bead*]! This [*pointing to action object*] made a red bead come out!" Which action produced a red or yellow bead was counterbalanced across children. After each demonstration, children were asked to select the correct action (via pointing [$n$ = 4] or by taking the action themselves [$n$ = 12]) to retrieve the bead ("Okay, I want a yellow bead. What should I do?"). If children responded incorrectly, the experimenter re-demonstrated the action and then prompted children again. One child was excluded for exceeding four demonstrations; all other children were able to respond correctly after fewer than four demonstrations ($M$ = 0.53 demonstrations, $SD$ = 0.80).

Next, children were introduced to the two stuffed animals, Lion and Monkey, one at a time and were told that each animal had a favorite color (red or yellow). The experimenter then asked children to retrieve a red or yellow bead for Lion or Monkey ("I want to get a yellow bead for Monkey to put in her cup. Do you think you can help me? What should I do?"). Children were asked to retrieve two beads for each animal. If children selected the wrong action, children were shown the correct action to produce the target bead and were given an additional attempt (capped at four re-demonstrations per object). One child was excluded for exceeding four incorrect responses for a single target; all other children were able to select the correct actions after fewer than four requests ($M$ = 0.84 requests, $SD$ = 1.02).

Whether children were introduced to Monkey or Lion first and the positions of Monkey and Lion on the left or right of the box were counterbalanced across children. By the end of the encoding phase, all children had successfully retrieved three beads of each color themselves. Fig. 1 (top right) shows a schematic of the encoding phase. The test phase began immediately following the encoding phase.

*Test phase.* On each test trial, the experimenter told children, "[Monkey/Lion] wants her favorite color bead. What should I do to get it for her? Can you show me?" (see Fig. 1, right panel). The experimenter never referred to the color by name, and children received no explicit feedback from the experimenter. If children selected the correct action, the experimenter released the correct color bead into the drawer, retrieved the bead, placed it in the relevant animal's cup, and said "Let's try another one!" If children selected an incorrect action, the experimenter said "Let's try another one!"

Children completed two blocks of 4 trials each, with order of the trials counterbalanced in each block, and children were counted as responding correctly if they selected only the action that would open the correct drawer.

### *Results*

We compared children's mean proportion correct responses across trials with chance (.50) using a one-sample $t$ test. We found that children were able to select the correct actions at rates significantly above chance ($M$ = .67, $SD$ = 0.25), $t(15) = 2.80$, $p = .01$, $d = 0.70$ (Fig. 2, left panel). Bayes factor analysis showed that the alternative hypothesis—that children performed at above-chance levels—was nearly four times more likely than the null hypothesis that children chose at chance ($BF_{10} = 3.88$).

A binomial logistic regression on children's responses with trial (1–8) as a factor and participant ID as a covariate revealed no main effect of trial ($\beta = .04$, $p = .63$). We also did not find a difference in performance when comparing mean performance on the first 4 trials with mean performance on the last 4 test trials, $t(15) = 0.57$, $p = .58$. Together, these results suggest that children's performance did not improve with repeated exposure to the test trials. The number of repeated demonstrations during encoding was slightly negatively correlated with test trial performance ($r = -.52$, $p = .04$), suggesting that children who took longer to learn the mappings during encoding performed worse at test than children who learned more quickly.

Because we counterbalanced the position of the animals across children, for half the children the position of the animals was adjacent to the actions that produced their favorite colors, and for the other half the position of the animals was opposite the actions that produced their favorite color beads. We asked whether children performed differently depending on the configuration of the animals. We found that children performed slightly better when the animal was adjacent to the action that produced their favorite color bead, but not significantly so ($M_{adjacent} = .78$, $SD = .26$; $M_{opposite} = .56$, $SD = .19$), $t(14) = 1.94$, $p = .073$.

Finally, we conducted an exploratory analysis to investigate whether children who chose correctly on the first test trial ultimately performed better overall. We ran an exploratory independent-samples $t$ test on children's mean proportion correct, comparing children who succeeded on the first test trial with those who did not. Results of this exploratory analysis suggested that children who succeeded on the first test trial ($n$ = 10; $M$ = .75, $SD$ = .25) did not perform significantly better overall than those who did not ($n$ = 6; $M$ = .54, $SD$ = .19), $t(14)$ = 1.76, $p$ = .101, $d$ = 0.94. Performance on the first trial did not predict performance on the other 7 test trials, $F(1, 14)$ = 0.49, $p$ = .49, $R^2$ = .03.

*Experiments 1 and 2 compared*

Experiments 1 and 2 used variations on a method to test the same hypothesis, namely that 2-year-olds could draw on brief experience at encoding to accomplish goals during test. Therefore, we asked whether children performed differently in Experiment 1 versus Experiment 2. Children performed similarly across the two experiments, $t(30) = -0.64$, $p$ = .53.

Next, we used the combined sample of Experiments 1 and 2 to examine whether children's ability to engage in memory-guided planning changed across our age range. We found that performance was positively correlated with children's age in months when controlling for experiment ($r$ = .58, $p$ = .001). Fig. 2 (right panel) shows each child's mean proportion correct as a function of their age for each experiment.

*Discussion*

In Experiment 2, we again asked whether 2-year-old children would be able to use episodic memories acquired over the course of an experiment to accomplish a goal-directed task, engaging in memory-guided planning. Children in Experiment 2 successfully chose the correct action to produce the target bead at above-chance levels, conceptually replicating the results of Experiment 1 while reducing task demands and controlling for a potential spatial confound. When combined, the results of Experiments 1 and 2 provide evidence that the ability to engage in memory-guided planning is present by the third year of life. Furthermore, this ability improved from 24 to 36 months of age, suggesting that memory-guided planning may emerge early during the third year of life and continue to develop across early childhood (see also Blankenship & Kibbe, 2019, for evidence from 3- and 4-year-olds).

## Experiment 3

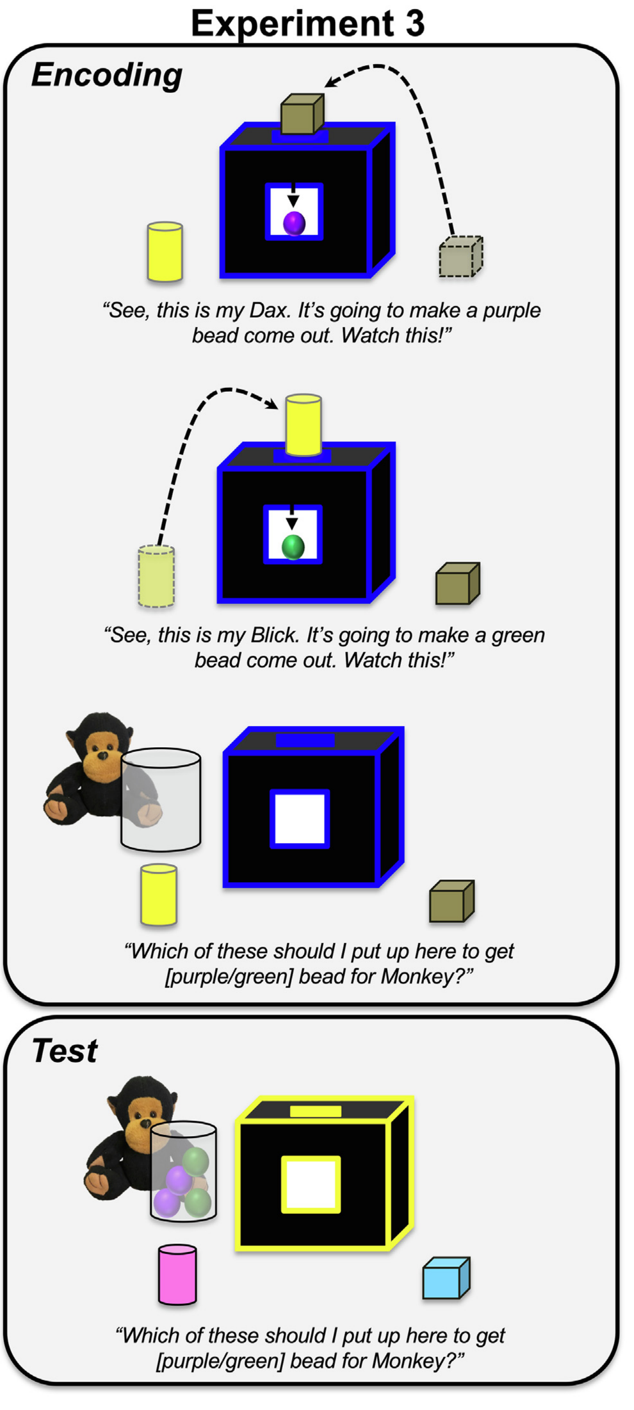
The ability to engage in memory-guided planning is a necessary component of episodic future-oriented thinking in which children draw on memories from past events to plan for possible future outcomes. Another necessary component of episodic future-oriented thinking is the ability to *generalize* from past experiences to novel contexts (e.g., Suddendorf et al., 2011). The results of Experiments 1 and 2 suggested that 2-year-olds can apply episodic memories to accomplish a novel goal within the same context that the memories were acquired. In Experiment 3, we investigated whether 2-year-olds may show some capacity for generalization in planning, flexibly applying episodes acquired during encoding to a completely novel context in order to accomplish a novel goal.

In Experiment 3, children were taught that placing different objects on top of a box would cause the box to release a specific colored bead. At test, we introduced children to similarly shaped, but perceptually distinct, bead box and objects, and children were asked to retrieve beads from the new box using the new objects. Critically, children never received any experience using the new box or objects. Previous work (Bauer & Dow, 1994) suggested that 16- to 20-month-olds are able to generalize functions of objects to similarly shaped novel objects when generating a learned sequence in a task that did not require planning. To succeed in our task, children needed to notice the similarities between the learned apparatus and the novel apparatus and then flexibly apply the episodic memories acquired during encoding to the new context in order to complete novel goals.

*Method*

*Participants*

A separate sample of 19 2-year-old children participated ($M$ = 30.98 months, range = 24.9–36.13; 12 girls). Parents identified their children as Asian ($n$ = 3), White ($n$ = 12, of whom 2 were identified as


Experiment 3
Encoding
"See, this is my Dax. It's going to make a purple bead come out. Watch this!"
"See, this is my Blick. It's going to make a green bead come out. Watch this!"
"Which of these should I put up here to get [purple/green] bead for Monkey?"
Test
"Which of these should I put up here to get [purple/green] bead for Monkey?"

Hispanic or Latinx), or multiracial ($n$ = 3) (1 parent did not report the child's race). Most parents self-reported that they had completed a college degree or higher. An additional 7 children were excluded because of experimenter error ($n$ = 1), refusal to participate in the task ($n$ = 4), exceeding four demonstrations during encoding ($n$ = 1), or failing to respond correctly to all prompts ($n$ = 1). We intended to include $N$ = 24 children in Experiment 3 but were forced to terminate data collection early due to the ongoing COVID-19 pandemic. A post hoc power analysis using the effect size obtained in Experiment 3 (one-sample $t$ test against chance, $d$ = 0.63) suggested a sample of $N$ = 18 was required to achieve 80% power with an alpha of .05. Children were recruited from the greater Boston area and received a small gift for their participation.

*Materials*

The stimuli included 20 2-cm colored beads (10 purple and 10 green), one stuffed animal (Monkey), four objects (i.e., two cubes and two cylinders) of roughly similar sizes (3–4 cm), one clear plastic cup, and two black foam core square boxes (18.5 × 16.5 × 16.5 cm). One of the boxes was outlined in blue tape and had a blue-outlined opening in its front (9.5 × 2.5 cm) with a blue square on top to indicate where to place objects; the other box was outlined in yellow tape and had a yellow-outlined drawer in its front with a yellow square on top (see Fig. 3).

*Procedure*

Children were seated at a small table with the experimenter seated in front of them and their caregiver seated behind them. Study sessions were recorded for later coding. The experiment was divided into *familiarization, encoding,* and *test* phases.

*Familiarization phase*

The purpose of the familiarization phase was to introduce children to the objects that would be used in both the encoding and test phases of the study. The experimenter told children that she was going to show them some of the objects that they would use in their "game" and then placed four objects (the yellow cylinder, the pink cylinder, the wooden cube, and the blue cube) on the otherwise empty table in front of children. She then drew children's attention to the similarities between the objects by saying, "See these objects? See how these are both square and these are both round?" The experimenter then removed the objects from the table.

*Encoding phase*

The experimenter placed the blue-outlined box on the otherwise empty table and told children, "This is my magic bead box! It makes purple and green beads. When I put something here [*pointing to small platform at the top of box*], a purple or green bead comes out here [*pointing to opening*]! She next placed the yellow cylinder and the wooden block on the table in front of the box (see Fig. 3) and proceeded to show children how to use the objects to retrieve green or purple beads (the order in which the objects were demonstrated was counterbalanced across participants). The experimenter told children, "I want a [green/purple] bead. I need to put one of these [*pointing to the objects*] up here [*pointing to top of box*] to make a [green/purple] bead come out. Hmm, which should I use?" She then lifted one of the objects and said, "I know! I'll use this [dax/blick]. See, this is my [dax/blick]. It makes [green/purple] beads come out. Watch this!" She then placed the object on the top of the box and released the relevant bead into the transparent drawer. She showed the bead to children and said, "Wow! The [dax/blick] made a [green/purple] bead come out!" The experimenter repeated this procedure for the second object. Thus, children were shown two object–label–outcome mappings: dax (wooden block)–green bead and blick (yellow cylinder)–purple bead.

**Fig. 3.** Schematic representations of the encoding and test phases in Experiment 3. The dashed arrows represent the movement of the action objects, and the solid arrow represents the movement of the bead following the placement of an action object on top of the box. The bottom panel shows an example of a test trial.

To gauge children's encoding of the object–label–outcome mappings, the experimenter asked children to point to the object that matched a given label (e.g., asking first "Which one is the dax?" and then "Which one is the blick?" with order of questions matching the order in which the objects were introduced to children). If children were unable to correctly identify which object matched a given label, the experimenter repeated the entire demonstration once again (with a maximum number of re-demonstrations set at two). After children had correctly identified each object ($n$ = 13) or reached the maximum criterion of two re-demonstrations ($n$ = 6), the experimenter moved on to the next phase of the experiment.

Children were next asked to retrieve beads from the box themselves. The experimenter said, "Okay, so now I want a [purple/green] bead. Which one these should I put up here to make a [purple/green] bead come out?" If children chose correctly, the experimenter placed the object on top of the box and released the relevant bead into the drawer. If children chose incorrectly, the experimenter placed the object on top of the box and said "Hmm, that one doesn't work!" and then prompted children to choose again. Children retrieved a total of four beads (either green, purple, purple, green or purple, green, green, purple; order counterbalanced across children).

Next, the experimenter introduced children to a single stuffed animal character, Monkey. Children were told that Monkey was collecting beads to put in her cup. The experimenter then indicated that Monkey wanted a certain colored bead and asked children for assistance ("Monkey wants a [green/ purple] bead to put in her cup. Which of these [*pointing to objects*] should I put up here [*pointing to top of box*] to get a [green/purple] bead for Monkey?"). Children again retrieved four beads (either green, purple, purple, green or purple, green, green, purple; whichever order they did not yet receive). If children chose correctly, the experimenter demonstrated that the object made the box produce the bead; if children chose incorrectly, the experimenter showed children that the object did not work and prompted them again.

Children were then again asked to indicate which object was the blick and which was the dax (asked in the opposite order from the first prompt). Of 19 children, 14 (74%) were able to correctly identify each object. The experimenter then asked children to indicate which object produced green beads and which produced purple beads (by pointing). Of 19 children, 13 (68%) were able to correctly identify which object produced which color beads.

By the end of the encoding phase, all children had successfully retrieved four beads of each color. The test phase began immediately following the encoding phase.

*Test phase: Generalization + planning*

The experimenter told children that Monkey wanted more beads but that the box they had been using was now empty. She removed the blue-outlined box from the table and placed the yellow-outlined box in its place. She then said, "Oh look! I found another magic bead box! This magic bead box also makes purple and green beads come out." She then placed the pink cylinder and the blue block on the table in front of the box (Fig. 3) and said, "Do you remember these objects? We can use these to make the beads come out, but I'm not sure which one makes purple and which one makes green beads. Do you think you can help me and Monkey?"

Children then completed 8 test trials in which they retrieved beads for Monkey (two blocks of 4 trials each [two green beads and two purple beads] with order of the trials counterbalanced in each block). On each trial, the experimenter asked children to retrieve either a green or purple bead for Monkey, saying "Monkey wants a [purple/green] bead. Let's think. Which one of these [*pointing to new objects on table*] should I put up here [*pointing to top of new box*] to get a [purple/green] bead for Monkey?"

Children received no explicit feedback from the experimenter during test trials. As in Experiment 2, if children selected the correct action, the experimenter released the correct color bead into the drawer, retrieved the bead, placed it in the Monkey's cup, and said "Let's try another one!" If children selected an incorrect action, the experimenter said, "Let's try another one!"

After completing the 4 test trials, the experimenter removed everything from the table except the pink cylinder and the blue block. To examine whether children had extended the labels to the novel objects, she asked children, "Which is the blick? And which is the dax?" (order counterbalanced across children). One child declined to respond to these questions. Of the 18 children who responded, 14 extended the labels to both new objects and 2 extended the labels to one of the new objects. The experimenter removed both objects from the table and then placed the yellow cylinder and wooden block on the table and asked children, "Which is the dax? And which is the blick?" Of 18 children who responded, 10 chose correctly for both objects and 4 responded correctly for at least one object.

*Results*

Children selected the correct object to produce the relevant bead at rates significantly above chance ($M$ = .64, $SD$ = 0.22), one-sample $t$ test against .50, $t(18)$ = 2.76, $p$ = .013, $d$ = 0.63 (Fig. 4), with Bayes factor showing that the alternative hypothesis was nearly four times more likely than the null hypothesis ($BF_{10}$ = 3.69). A binomial logistic regression on children's responses with trial (1–8) as a factor and participant ID as a covariate revealed no main effect of trial ($\beta = -.04$, $p$ = .67); children's performance did not change across trials. Children performed similarly on the first 4 test trials compared with the last 4 test trials, $t(18)$ = 0.60, $p$ = .56. Performance was not correlated with age ($r$ = .84, $p$ = .73). Fig. 4 (right panel) shows individual children's mean proportion correct responses as a function of age in months. The number of demonstrations children observed during the encoding phase also was not correlated with their performance at test ($r = -.17$, $p$ = .48).

Finally, we asked whether children who chose correctly on the first test trial, in which they encountered the novel objects and novel context for the first time, ultimately performed better overall than children who were not successful on the first trial. We ran an exploratory independent-samples $t$ test on children's mean proportion correct, comparing children who succeeded on the first test trial with those who did not. Results of this exploratory analysis suggested that children who succeeded on the first test trial ($n$ = 11; $M$ = .72, $SD$ = .25) performed significantly better overall than children who were unable to generalize on the first test trial ($n$ = 8; $M$ = .52, $SD$ = .10), $t(17)$ = 2.17, $p$ = .045, $d$ = 1.07. Performance on the first test trial did not reliably predict performance on subsequent trials, $F(1, 17)$ = 0.65, $p$ = .43, $R^2$ = .04.

*Discussion*

In Experiment 3, we asked whether 2-year-old children could apply episodic memories to complete a goal in a novel context. We found that children who learned the functions of two objects in one context were able to apply that knowledge to similar but distinct objects in a new context. This suggests that 2-year-olds may engage in memory-guided planning more flexibly, an important precursor to fully developed episodic future-oriented thinking.

Whereas the results of Experiments 1 and 2 suggested that children's ability to engage in memory-guided planning within the same context improved across the third year of life, children's success in Experiment 3 was not related to age. The ability to generalize episodes to new goal contexts may be an emerging skill in 2-year-olds and therefore may follow a more protracted developmental trajectory than executing plans within the same context that the relevant episodic memories were acquired. However, our results also suggested that there may be individual differences in the ability to generalize past episodes to new planning contexts that are not necessarily related to age. Children who generalized from past episodes to the novel context on the first trial performed better overall than children who failed in the first test trial (although these analyses were exploratory and therefore should be interpreted with caution). Therefore, success may be driven in part by the ability to notice the similarities between past episodes and novel context. We discuss these results and potential future directions next in the General Discussion.

## General discussion

The ability to apply past experiences to plan is critical for typical cognitive functioning and is one of the essential building blocks of future-oriented thinking, a uniquely human ability (Suddendorf & Corballis, 2007). Previous work investigated memory-guided planning starting at 3 years of age (Atance, 2015; Atance & O'Neill, 2005; Blankenship & Kibbe, 2019; Russell, Alexis, & Clayton, 2010;

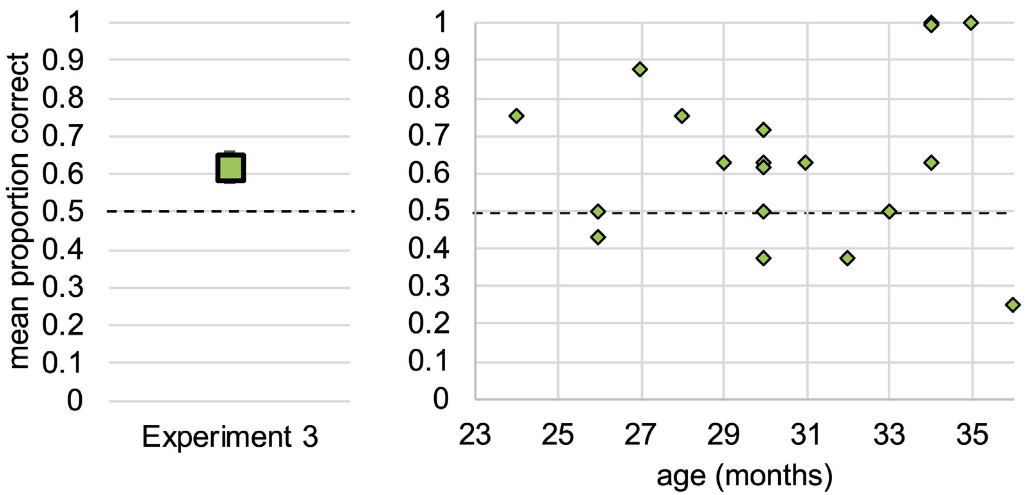



**Fig. 4.** Overall mean proportions correct (left panel) and individual children's mean proportions correct across trials as a function of age in months (right panel) in Experiment 3.

Suddendorf, et al., 2011) and found substantial developmental improvement from 3 to 6 years of age. We investigated the emergence of memory-guided planning by examining whether 2-year-olds could coordinate retrieval of episodic memories and execution of a plan to successfully engage in memory-guided planning both within the context that the relevant memories were acquired (Experiments 1 and 2) and in a novel context (Experiment 3). Our tasks were designed to be appropriate for 2-year-olds; the tasks made limited demands on children's productive language abilities, did not require verbal responses, and did not require children to use knowledge or memories acquired outside of the lab.

In Experiments 1 and 2, we asked whether children could selectively apply episodes acquired over the course of an experiment to accomplish goals. To succeed in our tasks, children needed to coordinate episodic memory (recalling the relevant episodes from the experiment), working memory (keeping track of the problem at hand), and planning (selecting and executing the relevant action). We consistently found that 2-year-old children were able to do so at rates significantly above chance. These results contrast with previous research suggesting that memory-guided planning abilities do not emerge until the fourth or fifth year of life (e.g., Atance & O'Neill, 2005; Russell et al., 2010).

The results of Experiments 1 and 2 also suggest that memory-guided planning abilities develop significantly across the third year of life. Memory-guided planning is a complex cognitive process, requiring coordination of memory encoding and retrieval and executive functions, all of which individually are undergoing significant development across the third year (Bauer & Dow, 1994; Bauer et al., 2003; Cheng, Káldy, & Blaser, 2020; Gagne & Saudino, 2016; Hendry, Jones, & Charman, 2016; Kibbe & Applin, 2021). Limitations to, and developmental change in, memory-guided planning could be driven by limitations and developmental change within these individual systems, developmental change in the ability to coordinate these systems to achieve a goal, or (likely) both. Although further work is needed to fully understand the mechanisms driving developmental change in memory-guided planning, the results of Experiments 1 and 2 illustrate the importance of including 2-year-olds in its developmental trajectory.

In Experiment 3, we asked whether 2-year-olds could apply episodic memories to a novel problem in a novel context, exhibiting flexibility in memory-guided planning. The task required children not only to coordinate memory and planning but also to generalize memories of the encoded apparatus to an apparatus with which they previously had no experience. We found that 2-year-olds were able to do so at rates above chance, suggesting that their memory-guided planning abilities are flexible. Interestingly, we did not observe developmental change in performance across our age range, suggesting that the ability to generalize episodic memories to novel contexts may be emerging in this age group and therefore may follow a more protracted developmental trajectory than memory-guided planning that does not require generalization. The ability to generalize from past episodes to novel contexts likely requires cognitive flexibility, a skill that is developing across the third year of life (Blakey, Visser, & Carroll, 2016; Deak, 2003; Kloo, Perner, Kerschhuber, Dabernig, & Aichhorn, 2008). We are currently investigating how individual differences in cognitive control may predict performance on memory-guided planning tasks that require generalization and on those that do not.

Although our results suggest that 2-year-olds have the ability to use memories to plan, this ability is still limited in many respects. For example, the ability to coordinate multiple episodes to engage in more complex plans develops significantly from 3 to 4 years of age (Blankenship & Kibbe, 2019). Although 2-year-olds are able to complete a plan using a single episodic memory, they may struggle when asked to complete plans involving more than one step or requiring integration of multiple event memories. Furthermore, although we found evidence that 2-year-olds can flexibly apply past episodes to a novel problem context, the novel problem was presented with no delay in time or displacement in space. Two-year-olds, whose episodic memory abilities are still developing, may struggle in more memory-demanding contexts (see, e.g., Liszkai-Peres et al., 2020). Further work is needed to examine sources of limitations and developmental change in memory-guided planning ability during early childhood.

Our results have implications for our understanding of the emergence and development of memory-guided planning and episodic prospection. One of the defining features of episodic prospection is the ability to generate projections for new events that have not yet happened, generalizing from past experiences to completely novel contexts at later time points (Atance & O'Neill, 2001; Suddendorf et al., 2011). We theorize that 2-year-olds' early-emerging memory-guided planning abilities, demonstrated here, may form one of the basic building blocks of fully developed episodic future-oriented thinking. As children develop, they may build on early memory-guided planning capacities to support more sophisticated future-oriented cognition.

## Acknowledgments

This research was supported by a grant (F32HD094554) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) awarded to Tashauna L. Blankenship. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health. We are grateful to the families for their participation in our research and to our research team members for their assistance with participant recruitment, data collection, and coding.

## References

Atance, C. M. (2015). Young children's thinking about the future. *Child Development Perspectives, 9*(3), 178–182.
Atance, C. M., Louw, A., & Clayton, N. S. (2015). Thinking ahead about where something is needed: New insights about episodic foresight in preschoolers. *Journal of Experimental Child Psychology, 129*, 98–109.
Atance, C. M., & Meltzoff, A. N. (2005). My future self: Young children's ability to anticipate and explain future states. *Cognitive Development, 20*(3), 341–361.
Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences, 5*(12), 533–539.
Atance, C. M., & O'Neill, D. K. (2005). Preschoolers' talk about future situations. *First Language, 25*(1), 5–18.
Bauer, P. J., & Dow, G. A. (1994). Episodic memory in 16- and 20-month-old children: Specifics are generalized but not forgotten. *Developmental Psychology, 30*(3), 403–417.
Bauer, P. J., Wiebe, S. A., Carver, L. J., Waters, J. M., & Nelson, C. A. (2003). Developments in long-term explicit memory late in the first year of life: Behavioral and electrophysiological indices. *Psychological Science, 14*(6), 629–635.
Blakey, E., Visser, I., & Carroll, D. J. (2016). Different executive functions support different kinds of cognitive flexibility: Evidence from 2-, 3-, and 4-year-olds. *Child Development, 87*(2), 513–526.
Blankenship, T. L., & Kibbe, M. M. (2019). Examining the limits of memory-guided planning in 3- and 4-year olds. *Cognitive Development, 52*, 100820. https://doi.org/10.1016/j.cogdev.2019.100820.
Busby, J., & Suddendorf, T. (2005). Recalling yesterday and predicting tomorrow. *Cognitive Development, 20*(3), 362–372.
Cheng, C., Káldy, Z., & Blaser, E. (2020). Coding of featural information in visual working memory in 2.5-year-old toddlers. *Cognitive Development, 55* 100892.
Cuevas, K., Rajan, V., Morasch, K. C., & Bell, M. A. (2015). Episodic memory and future thinking during early childhood: Linking the past and future. *Developmental Psychobiology, 57*(5), 552–565.
Deak, G. O. (2003). The development of cognitive flexibility and language abilities. In R. V. Kail (Ed.). *Advances in child development and behavior* (Vol. 31, pp. 271–327). San Diego: Academic Press.
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.
Gagne, J. R., & Saudino, K. J. (2016). The development of inhibitory control in early childhood: A twin study from 2–3 years. *Developmental Psychology, 52*(3), 391–399.
Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116*(2), 439–453.
Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*(6873), 755.
Gollwitzer, P., & Oettingen, G. (2011). Planning promotes goal striving. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 162–185). New York: Guilford.
Hayne, H., Gross, J., McNamee, S., Fitzgibbon, O., & Tustin, K. (2011). Episodic memory and episodic foresight in 3- and 5-year-old children. *Cognitive Development, 26*(4), 343–355.
Hayne, H., & Imuta, K. (2011). Episodic memory in 3- and 4-year-old children. *Developmental Psychobiology, 53*(3), 317–322.
Hendry, A., Jones, E. J. H., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review, 42*, 1–33.
Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.
Kibbe, M. M. (2015). Varieties of visual working memory representation in infancy and beyond. *Current Directions in Psychological Science, 24*(6), 433–439.
Kibbe, M. M. & Applin, J. B. (2021). Tracking what went where across toddlerhood: Feature–location binding in 2–3-year-olds' working memory. Manuscript submitted for publication.
Kloo, D., Perner, J., Kerschhuber, A., Dabernig, S., & Aichhorn, M. (2008). Sorting between dimensions: Conditions of cognitive flexibility in preschoolers. *Journal of Experimental Child Psychology, 100*(2), 115–134.
Liszkai-Peres, K., Kampis, D., & Király, I. (2020). The flexibility of early memories: Limited reevaluation of action steps in 2-year-old infants. *Journal of Experimental Child Psychology, 203* 105046.
Mau, W. C. (1995). Educational planning and academic achievement of middle school students: A racial and cultural comparison. *Journal of Counseling & Development, 73*, 518–526.
McCarty, M. E., Clifton, R. K., & Collard, R. R. (1999). Problem solving in infancy: The emergence of an action plan. *Developmental Psychology, 35*(4), 1091–1101.

Newcombe, N. S., Balcomb, F., Ferrara, K., Hansen, M., & Koski, J. (2014). Two rooms, two representations? Episodic-like memory in toddlers and preschoolers. *Developmental Science, 17*(5), 743–756.
Prabhakar, J., Coughlin, C., & Ghetti, S. (2016). The neurocognitive development of episodic prospection and its implications for academic achievement. *Mind, Brain, and Education, 10*, 196–206.
Prabhakar, J., & Hudson, J. A. (2014). The development of future thinking: Young children's ability to construct event sequences to achieve future goals. *Journal of Experimental Child Psychology, 127*, 95–109.
Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237.
Rovee-Collier, C. K., & Sullivan, M. W. (1980). Organization of infant memory. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 798–807.
Russell, J., Alexis, D., & Clayton, N. (2010). Episodic future thinking in 3- to 5-year-old children: The ability to think of what will be needed from a different point of view. *Cognition, 114*(1), 56–71.
Scarf, D., Boden, H., Labuschagne, L. G., Gross, J., & Hayne, H. (2017). "What" and "where" was when? Memory for the temporal order of episodic events in children. *Developmental Psychobiology, 59*(8), 1039–1045.
Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1481), 773–786.
Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic future thinking: Mechanisms and functions. *Current Opinion in Behavioral Sciences, 17*, 41–50.
Schwier, C., van Maanen, C., Carpenter, M., & Tomasello, M. (2006). Rational imitation in 12-month-old infants. *Infancy, 10*, 303–311.
Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science, 348*(6230), 91–94.
Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences, 30*(3), 299–313.
Suddendorf, T., Nielsen, M., & Von Gehlen, R. (2011). Children's capacity to remember a novel problem and to secure its future solution. *Developmental Science, 14*, 26–33.
Suddendorf, T., & Redshaw, J. (2013). The development of mental scenario building and episodic foresight. *Annals of the New York Academy of Sciences, 1296*(1), 135–153.
Xu, F., Cote, M., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science, 16*(5), 372–377.
Zhu, J., & Mok, M. M. C. (2012). Effect of academic goal orientation, goal setting, and planning on academic achievement of secondary students in Hong Kong. *Assessment and Learning, 1*(1), 11–30.