

*Developmental Psychology*

Repeated Sequential Action by Young Children:  
Developmental Changes in Representational Flexibility of Task Context

Kaichi Yanaoka and Satoru Saito

Address correspondence:

Kaichi Yanaoka

Graduate School of Education, Kyoto University,  
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

Tel. and Fax: +81-75-753-3004

E-mail. yanaoka.kaichi.22a@st.kyoto-u.ac.jp

© 2018, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI:  
10.1037/dev0000678

### Abstract

A wealth of developmental research suggests that preschoolers are capable of reporting, imitating, and performing sequential actions they engage in routinely. However, few studies have explored the developmental and cognitive mechanisms required for learning how to perform such routines. A previous computational model of routines argued that a representation of task contexts underlying routines could change flexibly. This position was supported by the empirical evidence that if adults are interrupted in the course of a routine, they make fewer errors if they are interrupted just before the selection of context-dependent action than if they are interrupted earlier. Another computational model examined how efficiently adults learned to perform routines and suggested the relationship of the learning efficacy with executive functions. The present study aimed to examine whether the above-mentioned models and evidence from adults can be extended to preschoolers by using an experimental task, in which children were required to play the role of a baker and repeatedly make toast for either a cat or mouse, with momentary distractions. Experiment 1 showed that earlier interruption tended to cause older children to produce more branch point errors than interruption immediately before the branch points, whereas younger children tended to be vulnerable to both interruptions. Further, across two experiments, this study showed that the developmental differences in how young children represent task contexts were associated with their executive functions. These findings indicate that the representational flexibility of task contexts underlies children's performance of repeated sequential actions and its association with executive functions.

Keywords: routine, task context, action development, executive functions, preschoolers

## Repeated Sequential Action by Young Children: Developmental Changes in Representational Flexibility of Task Context

### 1. Introduction

Our daily life is composed of several sequential actions that we perform routinely, such as making breakfast, taking a train, and eating out at a restaurant. Previous studies have successfully modelled how healthy adults and adults with neurological impairments learn such sequential actions (e.g., Arnold, Wing, & Rotshtein, 2017; Humphreys, Forde, & Francis, 2000; Ruh, Cooper, & Mareschal, 2010). While some of these studies have proposed good accounts of learning routines (Botvinick & Plaut, 2004; Cooper, Ruh, & Mareschal, 2014), they do not cover the development of sequential actions among preschoolers, which is the target of the current study.

From a developmental perspective, three- to four-year-olds are capable of reporting what happens in familiar, frequently experienced events (e.g., Fivush, 1984; Nelson, 1986). Moreover, previous studies (Bauer & Hertgaard, 1993; Bauer & Mandler, 1989, 1992; Bauer & Thal, 1990; Freier, Cooper, & Mareschal, 2015; Loucks & Meltzoff, 2013) have demonstrated that not only three-year-olds but also toddlers can accurately imitate sequential actions they perform routinely (e.g., putting a bear in the tub, making a sandwich). Furthermore, preschoolers are able to accurately perform such routines without explicitly being instructed how to perform them (Hudson & Fivush, 1991; Hudson, Shapiro, & Sosa, 1995; Shapiro & Hudson, 2004; Yanaoka, 2014; Yanaoka & Saito, 2017). Although previous studies examine such competence in preschoolers, they have neglected how these children learn to perform routines. It is worthwhile to investigate whether the above-mentioned models (i.e., Botvinick & Plaut, 2004) and evidence from adults (i.e., Botvinick & Blysm, 2005) can be extended to preschoolers to explain how

they learn to perform routines. The objective of this study, therefore, was to clarify the developmental mechanism underlying the acquisition of routines in children. More specifically, we aimed to examine whether and how preschoolers differ from adults in the way they learn to perform sequential actions as routines.

### 1.1. Existing Models for Routines

A significant number of studies have indicated that the performance of routines requires the involvement of broader task context over the course of the sequential actions (Botvinick & Blysmá, 2005; Botvinick & Plaut, 2004; Cooper & Shallice, 2006; Lashly, 1951; Miller, Galanter, & Pribram, 1960; Schank & Abelson, 1977). Within the literature, therefore, many accounts have focused on the task contexts underlying sequential actions and explored how individuals learn to perform routines based on context information (e.g., Altmann & Trafton, 2002; Botvinick & Plaut, 2004; Cooper & Shallice, 2006).

In particular, Botvinick and Plaut (2004) proposed an influential model for routines (for a review, see Botvinick, 2008). Their model takes the form of a simple recurrent network that maps perceptual input units to action output units via a set of internal hidden units. Its recurrent connectivity enables the network to preserve and transform the context representations that are distributed across units. Botvinick and Plaut (2004) applied this network to the sequential actions used in coffee-making. Their model accounted for error-free performance of coffee-making, action slips under conditions of distraction, and the disorganization of action that occurs in adults with neurological impairments. One of the critical aspects of Botvinick and Plaut's (2004) model is that the simple recurrent network learns how context information is represented depending on its relevance to action selection (Botvinick & Plaut, 2002). For example, although the sequential actions for making coffee and tea overlap in many ways, some actions are context dependent: we

might habitually add cream to our coffee but not to our tea, for instance. To choose the context-dependent action, we need to remember the context or goal either implicitly or explicitly: are we making coffee or tea? Thus just before the context-dependent action (i.e., at a branch point in two similar but different sequential actions), we represent the context robustly because it matters at that point; whereas context representations are less influential earlier in the sequence, at the point(s) when they play no role in determining our actions. According to Botvinick and Plaut (2004), context representations, which are assumed to be implicitly represented, change flexibly over the time course of sequential actions, based on whether the contextual information is actually needed for action selection.

## 1.2 Representational Flexibility of Task Contexts in Preschoolers

To explore the developmental mechanism underlying the acquisition of sequential actions, we investigated whether the model of Botvinick and Plaut (2004) could be applied to a developmental sample. More specifically, we tested whether context representations would change flexibly while preschoolers performed a sequential action repeatedly.

Botvinick and Blyma (2005) conducted the first experimental study to test how context representations could change flexibly during the repeated execution of a sequential action; they examined the frequency of action slips generated by adult participants after brief interruptions during a sequential action that consisted of several subtasks performed sequentially. The sequential action was to prepare a cup of coffee; thus, the subtasks were, for example, to add coffee grounds, to add sugar, and to add cream, and so on. Each subtask consisted of several subactions, for example, open a packet of sugar, stir sugar in, throw away the packet. The selection of the subtask actions is context dependent, that is, for example, if sugar was added already, the next subtask should not be to add sugar. Participants were required to prepare 50 cups of coffee

in total. Brief interruptions were inserted during the second or third subtask. The position of the interruption within the subtask was either at the middle (e.g., while stirring in) or the end of the subtask (e.g., while holding a packet before throwing it away). Results showed that the middle subtask interruption led to more errors in selecting the context-dependent action for the next subtask than the end subtask interruption did. This finding suggests that context representations become attenuated at the middle of the subtask sequences but recover at the end of the subtask, and that interruption at the point when context representations are attenuated or degraded results in action selection errors.

In the current study, based on the coffee-making task used in Botvinick and Blysmá (2005), we developed a new task that required preschoolers to perform novel sequential actions repeatedly. In this task, children were asked to make toast for either a cat or a mouse. This means that there were two task contexts; the cat context and the mouse context. Importantly, some common subtasks (e.g., spreading butter on toast) were performed across the cat and mouse contexts, while one unique subtask (e.g., putting cheese on the toast) after performance of the common subtasks was specific to the mouse context. The point where the transition occurs between common subtasks and the mouse-unique subtask is called a *branch point*. The aim of this subtask arrangement was to investigate whether children's errors included lapses from one task context into another, a form of error often observed in human behaviour (Botvinick & Plaut, 2002). The intrusion from the context of the mouse into that of the cat, for example, a child executing the mouse-unique subtask in the context of the cat, could occur at the branch point. One might expect that the representations of cat and mouse task contexts would become degraded at the middle of a common subtask, because the context representations are not actually relevant to executing the common task. However, the task contexts were assumed to recover and

to be more saliently represented at the end of the common subtask, where the context representations are necessary for selecting the mouse-unique subtask. To examine whether the result of Botvinick and Blysmá (2005) could be extended to preschoolers, we compared rates of branch point errors caused by momentary disruption occurring at the middle of and at the end of a subtask that occurred just before a branch point.

### 1.3 Developmental Differences in How Preschoolers Learn to Perform Routines

In this study, we also investigated developmental changes in how task contexts were represented in preschoolers. During the preschool years, it has been reported, children experience developmental changes in many aspects of their cognitive abilities, such as theory of mind, executive functions, and math achievement (e.g., Bonny & Lourenco, 2013; Carlson & Moses, 2001; Munakata, Snyder, & Chatham, 2012). In the domain of sequential actions that we perform routinely, a recent study (Freier et al., 2015) suggested the course of developmental changes in children's performance of routines. They demonstrated that although both three- and five-year-olds accurately imitated the sequence of making a sandwich, the introduction of a misleading demonstration of sandwich-making that included irrelevant actions (e.g., peeling a banana) caused the three-year-olds to overimitate the irrelevant actions, while was not the case in the five-year-olds. Furthermore, offering spatial cues (i.e., arranging relevant items in the appropriate order) helped the three-year-olds to avoid overimitating the irrelevant actions. This study suggests that older children, similar to adults, learn and perform sequential actions in a goal-directed manner, but that younger children rely more on perceptual information rather than the goal and its associated context. This further indicates that younger children might not be able to modulate context representation flexibly according to the task goal. In relation to the current research paradigm, one may infer that older children are more likely to behave like adults, who

represent context representations robustly just before a branch point. Therefore, it is predicted that we will replicate the findings from Botvinick and Blysmá (2005) in older children but that this will not be the case for younger children, who cannot modulate context representation over the course of the sequential action.

#### 1.4 Executive Functions and Learning to Perform Routines

It is also important to determine what ability underlies the development of performance of sequential actions in a routine situation. According to Botvinick and Plaut (2004), our ability to learn to perform routines depends on the number of repetitions or frequency of task sequences. Thus, if we were to find developmental differences between older and younger children, even though children in both age groups experience the same number of trials, the model by Botvinick and Plaut (2004) would not adequately explain the developmental differences.

A recent model updated the model of Botvinick and Plaut (2004) by adding a unit of explicit goal representation to the network, furnishing a hierarchical structure (for a review, see Cooper et al., 2014). The updated version of the model takes a representation of the current goal as input, generating an action and a representation of the predicted goal. The goal representation was modulated by an executive control system and fed back into the model in the next processing step. The model was not only able to replicate the findings of Botvinick and Plaut (2004) using goal-based learning, but also to elicit more efficient learning of complex sequential actions than in their model. Given that the executive control system is constrained by individuals' working memory capacity (e.g., Shallice & Burgess, 1996) and by developmental factors (Munakata et al., 2012), these findings suggest that the efficiency of learning to perform routines varies depending on individual and developmental differences in the ability to manage goal representations or in executive functions.



In the present study, following the model by Cooper et al. (2014), we focused on the development of an executive control system, which it is assumed modulates how we perform sequential actions through executive functions. It is well known that executive functions, that is, the processes involved in the conscious control of thought and goal-directed behaviour (e.g., Miyake et al., 2000), dramatically improve over the preschool period (Best & Miller, 2010; Blair & Razza, 2007; Garon, Bryson, & Smith, 2008; Moriguchi, 2014; Zelazo et al., 2003). Moreover, Yanaoka and Saito (2017) demonstrated that five-year-olds progressively develop their ability to control sequential actions based on the maintenance of goal representations and that this ability is related to their executive functions. In our study, we examined the relationship between executive functions and how preschoolers learn to perform routines.

## **2. Experiment 1**

The first aim of Experiment 1 was to explore how preschoolers represented task contexts in executing sequential actions repeatedly. Extending the work of Botvinick and Blysmá (2005), preschoolers were required to repeatedly make toast for either a cat or a mouse, with momentary distractions. These interruptions, in which they were instructed to perform a digit span task, were inserted either at the middle or the end of the subtask, immediately prior to a branch point. We evaluated the impact of these two types of interruptions on the rate of branch point errors. Our prediction was that older children would produce more branch point errors due to the middle subtask interruption than due to the end one, whereas younger children would be susceptible to both types of interruptions.

Our second aim in Experiment 1 was to identify the relationships between well-established executive function tasks and the mechanism by which preschoolers learn to perform routines. Based on the goal circuit model (Cooper et al., 2014), we predicted that children with

high executive functions would exhibit a higher rate of branch point errors in the trials involving middle subtask interruption than in the trials involving end subtask interruption, whereas children with low executive functions might show higher levels of branch point errors in the two interruption conditions.

## 2.1 Method

### 2.1.1 Participants

Participants in Experiment 1 included 43 children (19 boys and 24 girls) attending a kindergarten school in Japan ('kindergarten' in Japan is equivalent to preschool or nursery school in many countries). Of the children who were tested, five-year-old child ( $n = 1$ ) and four-year-old children ( $n = 2$ ) were excluded from data analysis due to their failure to understand the experimental setting or cooperate. The final sample consisted of 20 four-year-old children ( $M = 51.22$  months,  $SD = 4.05$  months, age range: 47–58 months) and 20 five-year-old children ( $M = 65.21$  months,  $SD = 3.63$  months, age range: 60–70 months). None of the participants had any history of neurological disorders or neurodevelopmental delay, and all of them were native Japanese speakers. Their socioeconomic background was predominantly middle class. Informed consent was obtained from the parents or the kindergarten staff members for all children prior to their participation. This and subsequent studies were approved by the institutional ethics committee for experimental psychology research at Graduate School of Education, Kyoto University (approval number: CPE-162; title: 'The mechanism for representing task contexts during the execution of scripts').

### 2.1.2. Procedure

The individualized experiment was divided across the two days. On both days, children performed a toast-making task, wherein they prepared two kinds of toast while frequently being

interrupted. In addition, they performed the following four tests, which are often used for assessing executive functions: a red-blue task (Simpson & Riggs, 2005, revised), the standard Dimensional Change Card Sort (DCCS) task (Zelazo, 2006), the advanced DCCS task (Chevalier & Blaye, 2009), and a nine box task (Diamond, Prevor, Callender, & Druin, 1997, revised). The two versions of the DCCS tasks were conducted on the same day, and the other two tests were conducted on another day. The order of executive function tasks was counterbalanced across participants. To put the children at ease, both sessions were conducted in a quiet room at the kindergarten. The experimenter visited the kindergarten for several days prior to the experiment to establish rapport with the children. In both sessions, the testing required approximately thirty minutes.

### 2.1.3 Measures

#### 2.1.3.1 Red-blue task

This task was conducted to measure the children's inhibition abilities. The procedure was essentially the same as that of the black-white task used by Simpson and Riggs (2005). At the beginning of this task, the children were asked to say "blue" when the experimenter held up a blue card and "red" when a red one was held up. All the participants were able to answer correctly. Then, before the test phase, the children were instructed to point to the red card when experimenter said "blue" and point to the blue card when said "red". If they answered correctly in a practice trial of this opposite-color pointing, the experimenter praised them and continued to engage in the practice trial with the other color. If they answered incorrectly, the task instruction was repeated. This procedure was repeated until each child responded correctly to two practice trials in succession. In the test phase, children engaged in 16 trials, during which the

experimenter said “red” and “blue” eight times, each in a pseudo-random sequence. The total score consisted of the number of correct trials during a session (0–16).

#### 2.1.3.2 Standard DCCS

This task was conducted to measure shifting abilities. The procedure closely followed the protocol described by Zelazo (2006). Children were presented with two trays and two target cards. The experimenter labelled the target cards using two dimensions (yellow and green, cup and ship). In the shape (color) game, children were instructed to place all the cup cards (all the yellow cards) in the box with the yellow cup, and to place all the ship cards (all the green cards) in the box with the green ship. For each trial, the experimenter presented a test card and asked the participants to sort the cards according to the relevant rule. Six test cards were presented in a pseudo-random sequence, and no feedback was provided. Once the pre-switch phase was completed, the experimenter provided the dimension switch instruction, that is, introduced the children to a new game in which they were instructed to sort the test cards according to the other dimension. In the post-switch phase, the children were presented with six test cards and were asked to sort them according to the new rule. The rule was counterbalanced across participants. The total score represented the number of correct trials only during the post-switch phase.

#### 2.1.3.3 Advanced DCCS

This task was also administered to measure shifting abilities. The advanced DCCS task is more difficult to perform correctly than the standard DCCS because rule switches are more frequent. After completing the standard DCCS, the children were instructed to play the card-sorting game, but with a new rule: for each trial, the experimenter explicitly told children which dimension (color or shape) they were required to sort the test card based on. In contrast to Zelazo (2006) who used a visual cue (e.g., a card with a star), we used a verbal cue, as the visual cue

was too demanding for four-year-olds to set goal representations (Chevalier & Blaye, 2009). Each participant underwent 10 trials, including 5 color trials and 5 shape trials; their scores represented the number of accurate sortings.

#### 2.1.3.4 Nine Box Task

The nine box task in this study, an adapted version of the task used by Diamond et al. (1997), was conducted to measure updating abilities. The procedure for this task was very similar to the one used by Wiebe et al. (2011). At the start of the trials, the children were presented with nine boxes with lids of different patterns and colors. The experimenter placed a marble in each box, then closed all nine lids. The children were instructed to remember which box contained a marble, without receiving any information regarding its placement in any of the boxes. The children then identified the placement of the marbles hidden in all the boxes. They were allowed to reach for any box in any order, but they could open only one box during a trial. After they opened the box, the children received feedback (“You have found it” or “There is nothing in there”), and the box was then closed. A delay period of five seconds began from the moment the opened box was closed. During this period, the children closed their eyes for five seconds, while the experimenter scrambled all the boxes randomly. This procedure was repeated until they found all the marbles; the maximum number of trials was sixteen. The score was measured as the proportion of correct responses.

#### 2.1.4 Toast-Making Task

##### 2.1.4.1 Materials

In order to reproduce the procedure of making toast, we prepared toy bread, cream-colored clay, and yellow felt, which acted as substitutes for toast, butter, and cheese, respectively. Each ingredient was placed in a white plastic box. The boxes were arranged in a specific

sequence on a desk to help children take the ingredients in the appropriate order: toast (left), butter (center), and cheese (right). In addition to the ingredients, a toaster oven made of cardboard was also prepared, and placed next to the box of cheese (see Figure 1). In the trials with interruptions, a laptop computer (Microsoft Surface Pro 4) was set up on an adjacent desk so that children would hear the sound of a telephone ringing in the middle of making toast. After children touched an image of a telephone on the computer screen to stop the sound, an auditory cue of three digits between 1 and 9 was presented three times through the computer, at a rate of 1 per second; no digit could appear twice in the same trial. The children were then asked to recall the digits.

-----  
Insert Figure 1 about here  
-----

#### 2.1.4.2 Procedure

First, the participants were instructed to play the role of baker and make toast for either a “mouse” or a “cat.” The experimenter explained to the children that toy bread, cream-colored clay, and yellow felt symbolized toast, butter, and cheese, respectively, for making toast (see Figure 1). They were also introduced to different toast recipes, for a mouse and a cat: the recipe for the mouse contained both butter and cheese because it liked both, but the recipe for the cat contained only butter, because the cat liked butter but not cheese. The experimenter also demonstrated how to make toast for the mouse and cat. For the mouse, the experimenter picked a plate, placed four pieces of toy bread on the plate, spread yellow clay on the toast, put yellow felt on the clay, placed the toast in the toaster oven, and set the completed toast on the adjacent desk.

The process of making toast for the cat was similar to that for the mouse, with the exception that cheese was not added. After the demonstration, the experimenter asked the children to verbally describe how they would make toast for both the cat and the mouse, to confirm that they had understood the instructions. All the children except for 1 four-year-old child were able to answer the question.

In two-thirds of the trials, the toast-making task was interrupted. In this condition with interruptions, we used the same procedure as Botvinick and Blyma (2005); each trial was interrupted either at the middle or the end of spreading butter, which was followed by a branch point—the point of putting cheese on the toast for the mouse, but not for the cat. The middle interruption occurred immediately after spreading yellow clay on the two pieces of toast, whereas the end interruption occurred as children finished spreading yellow clay on all four pieces of toast. Before the trials had begun, children were instructed that as soon as the phone rang via computer, they were to stop what they were doing, put down all the objects they were holding, and answer the phone. Children were then required to recall the three lists composed of the three digits, in the correct order. Immediately afterward, they were instructed to resume making toast.

The participants engaged in four practice trials (two mice and two cat trials); half of the trials were interrupted. At the beginning of each trial, children were presented with the picture of a cat or a mouse and encouraged to make toast for the presented animal. To prevent children from checking task contexts again while making toast, the pictures were hidden from them. If they made mistakes in executing the action sequences, they were provided feedback by the experimenter. Afterward, they performed 12 test trials per day. After 6 trials, the children were given a short break, during which they received stickers as a “salary.” Across the two days, all

sessions consisted of 4 mouse trials and 4 cat trials for each condition: the control, middle interruption, and end interruption conditions.

#### 2.1.4.3 Scoring

Our index was the rate of branch point errors across the test trials. Branch point errors occurred at the transition point from the subtask of spreading butter to the next subtask of adding cheese. In this study, the branch point errors reflected the intrusion of one context into another, resulting in context-inappropriate action. Two specific patterns of branch point errors were possible: reaching toward the cheese, in the context of making toast for the cat, and placing the toast in the oven without putting cheese on it, in the context of the mouse. In some cases, children reached toward an incorrect target or stopped reaching, and they were redirected toward the correct target. Schwartz, Reed, Montgomery, Palmer, and Mayer (1991) referred to such errors as ‘microslips’. Following the scoring used in Botvinick and Blysmá (2005), to maximize the sensitivity of error detection, microslips were included in the branch point errors.

All the trials were coded by the experimenter and a research assistant, an undergraduate psychology major, based on a video-recording. The two coders identified 22 trials that involved imprecise time of interruptions; such trials were excluded from later analysis. The remaining trials were evaluated for the occurrence of branch point errors. The consistency rate was 91% ( $\kappa = .86$ ) between the two coders, and coding disagreements were resolved through discussion.

## 2.2 Experiment 1 Results and Discussion

### 2.2.1 Data Analysis

In the toast-making task, a trial was scored as incorrect if the children made a branch point error in the trial. Analyses of variance are not appropriate for dichotomous response variables; thus, we analysed the data with generalized linear mixed models (GLMMs) using the



lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in the R system for statistical computing (R Core Team, 2013). GLMMs with binominal error distribution were used to account for nonindependence in the binary data (i.e., repeated measurements per child; Baayen, 2008). We conducted two sets of analyses. First, we sought to identify whether rates of branch point errors in the two types of interrupted trials (i.e., the interrupted conditions)<sup>1</sup> were higher than those in the uninterrupted trials (i.e., control condition), which would suggest that the interruptions led to an increase in error rates. Second, to test our key prediction, error rates of branch points were compared between trials in the middle interruption condition and in the end interruption condition; this comparison would indicate which time point (middle or end) was more susceptible to the interruption. Thus, we conducted planned comparisons and adapted Helmert coding<sup>2</sup> using an interruption timing factor (control, middle, end). To compare the control condition with the average of the other two interrupted conditions (hereafter, the *interruption factor*), we coded the control condition as 0.67, the middle interruption condition as -0.33, and the end interruption condition as -0.33. In another comparison of the middle interruption condition with the end interruption condition (hereafter, the *timing factor*), we coded the control condition as 0, the middle interruption condition as -0.5, and the end interruption condition as 0.5. The following variables were coded to be centred: animal (mouse = 1, cat = -1), day (day 1 = -1, day 2 = 1), and sex (boy = 1, girl = -1).

Preliminary analysis indicated no main effect of day ( $b = -0.03, z = -0.38, \chi^2 = 0.12, df = 1, p = .723$ ) or of sex ( $b = -0.18, z = -1.45, \chi^2 = 2.06, df = 1, p = .151$ ), indicating that there were

---

<sup>1</sup> Regarding the performance on the secondary task, all the participants could answer all the lists correctly.

<sup>2</sup> Contrast coding provides a series of precise comparisons among the group means. Helmert coding is one kind of contrast coding, and compares each level of a categorical variable to the mean of the subsequent levels.

no differences in performance between Day 1 and Day 2 or between boys and girls; thus, the day and sex factors were excluded from subsequent analyses. Our first analysis explored age differences in the extent to which both interruption and timing factors affected the rates of branch point errors. Further, we compared the additional extent to which executive function tasks moderated the effects of both interruption and timing factors on the rates of branch point errors, beyond age differences.

To identify the most parsimonious model to explain the variance of rates of branch point errors, all possible models were compared using the Akaike's Information Criterion (AIC) and excluded variables that did not contribute to decreasing AIC of the model. Fixed effects were also tested on the basis of whether their inclusion improved the model fit using likelihood ratio tests (Barr, Levy, Scheepers, & Tily, 2013).

### 2.2.2 Developmental Changes of Routines

First, based on the findings from adults (Botvinick & Blyma, 2005), we aimed to examine whether preschoolers produced more branch point errors in the middle interruption condition than in the end interruption condition. Thus, our analysis included the factor of age, the interruption factor, the timing factor, and their two-way interactions (age x interruption factor, age x timing factor) as fixed-effect variables. For the random effects, a random intercept for participant and animal, and a random by-participant slope for the interruption factor were included.

The best-fitting model (AIC = 869.3), which fits significantly better than the null model (AIC = 889.3,  $p < .001$ ), is shown in Table 1. The results of this mixed logit model revealed a significant main effect of the interruption factor. This result indicated that the rate of a branch point error in the interrupted trials was higher than in the trials of the control condition. There

was also a significant main effect of the timing factor, reflecting that the rate of a branch point error in the middle interruption condition was higher than in the end interruption condition. We found a trend of significant interaction between age and the timing factor. A simple slope test revealed that the main effect of the timing factor was not significant in younger children ( $b = 0.26, z = 0.91, p = .362$ ), whereas the main effect of the timing factor was significant in older children ( $b = 1.00, z = 3.39, p < .001$ ).<sup>3</sup> Overall, these findings revealed that the previous findings for adults (Botvinick & Blysm, 2005) could be extended to preschoolers as well. The interaction between age and the interruption factor was not significant ( $\chi^2 = 0.03, df = 1, p = .864$ ), but it was suggested that there might be developmental differences in how children represented task contexts, although minor age differences were found in performance on the control condition.

-----  
Insert Table 1 about here  
-----

### 2.2.3 Executive Functions and Developmental Changes of Routines

In the current study, we also aimed to explore whether executive functions moderated the effect of either or both the interruption and timing factors on the rates of branch point errors, beyond the factor of age; thus, we added the factors of the four executive function tasks to the above model to explore age differences. As shown in Table 2, a correlation analysis was conducted for checking the relationships among four executive functions tasks and their association with the performance of the toast-making task in each condition. The advanced

---

<sup>3</sup> Simple slopes between age and timing factor were calculated at 1 standard deviation above the mean of age (older children) and at 1 standard deviation below the mean of age (younger children).

DCCS task was significantly related to the red/blue task and the nine box task. Importantly, there was a significant relationship between the advanced DCCS and the correct rate of the toast-making task only for the end-interruption condition. These results suggested that the advanced DCCS task was the most sensitive measure among the four executive tasks and closely related to overcoming the end subtask interruption; therefore we reported a model including interactions related to the advanced DCCS task. Thus, our analysis included main effects of age, four executive function tasks, the interruption factor, the timing factor, and two-way interactions (advanced DCCS x interruption factor, advanced DCCS x timing factor) as fixed-effect variables. The random effect variables consisted of a random intercept for participant and animal, and a random by-participant slope for the interruption factor.

-----  
Insert Table 2 about here  
-----

The results of the best-fitting model (AIC = 859.6) are shown in Table 3. We compared these two models using a likelihood test, revealing that the model including the executive function tasks (Table 3) fitted significantly better than the model exploring only the age differences (Table 1) ( $\chi^2 = 15.74$ ,  $df = 3$ ,  $p = .001$ ). We found a significant main effect of the red-blue task, indicating that children with a high score on the red-blue task produced fewer branch point errors than children with a low score on the red-blue task. Moreover, a significant interaction between the interruption factor and the advanced DCCS was identified (see Figure 2). A simple slope test revealed that children who scored low on the advanced DCCS produced more branch point errors in the two interruption conditions than in the control condition ( $b = 1.27$ ,  $z = 4.18$ ,  $p < .001$ ). However, those who scored high on the advanced DCCS did not

produce more branch point errors ( $b = 0.34, z = 1.08, p = .320$ ). Next, we identified the significant interaction between the timing factor and the advanced DCCS (again see Figure 2). The positive coefficient indicated that for children with high performance on the advanced DCCS, the rate of branch point errors was higher in the middle interruption condition than in the end interruption condition ( $b = 1.24, z = 3.42, p < .001$ ), but that significant differences of the error rates between the two interruption conditions were not observed in children with low performance on the advanced DCCS ( $b = 0.23, z = 0.84, p = .402$ ).

-----

Insert Table 3 about here

-----

-----

Insert Figure 2 about here

-----

Overall, these findings suggest that executive functions are required to overcome momentary interruption in performing routine action sequences, regardless of the timing of the interruptions. In addition, our findings clearly demonstrated that children who scored high on the advanced DCCS, similar to adults, showed fewer branch point errors in the end interruption condition than in the middle interruption condition, while those scored low showed many branch point errors in both interruption conditions, indicating that executive functions measured by advanced DCCS might modulate context representations for the toast-making task.

### 3. Experiment 2

Experiment 1 provided the first evidence of developmental change in assumed context representation for the learning and execution of sequential actions and its relationship with the development of executive functions. Experiment 2 aimed to validate the findings of Experiment 1. We used nearly the same procedure as the toast-making task in Experiment 1, with the exception of three points. First, in Experiment 2 the toast-making task was conducted in only one day, because the factor of day had had little impact on the children's performance in Experiment 1. Second, we expanded the age range to also include a slightly older group (six-year-olds). Executive functions in six-year-old children should be more developed than those in four- and five-year-olds, and the performance of executive function tasks in a wider range of age groups will have a larger variation. It was expected that we could replicate the effects of executive functions on susceptibility to two kinds of interruption with a wider range of ages of children. Third, we added one more common subtask to the recipe for making toast. Ruh et al. (2010) demonstrated that a longer subtask induced longer response time for selecting an action at a branch point. Although fewer trials were conducted in Experiment 2 (one-day experiment) than in Experiment 1 (two-day experiment), longer duration in conducting common subtasks might lead to more branch point errors and enable us to detect results similar to those in Experiment 1.

Overall, we aimed to validate the findings reported in Experiment 1 and extend them to a broader range of age groups and to the case of longer sequential action.

### 3.1 Methods

#### 3.1.1 Participants

To identify the minimum sample size for sufficient statistical power, we carried out Monte Carlo simulations using the `powerSim` function from the `simr` package in R (Green & Macleod, 2016). The simulation revealed that a sample of  $N = 72$  participants yielded a power of

$\beta = 0.82$  (95% CI = 0.80, 0.85) to detect medium fixed-effects of size 0.5 for the interaction between the timing factor and the advanced DCCS, which was our main interest, at an  $\alpha$  level of .05. A total of 77 children (37 boys and 40 girls) attending kindergarten schools in Japan participated in Experiment 2, but six-year-old ( $n = 1$ ), five-year-old ( $n = 1$ ), and four-year-olds ( $n = 3$ ) were excluded due to failure to cooperate. Thus, 23 four-year-old children ( $M = 55.5$  months,  $SD = 3.26$  months, age range: 49–60 months), 24 five-year-old children ( $M = 65.4$  months,  $SD = 3.99$  months, age range: 61–72 months), and 25 six-year-old children ( $M = 76.8$  months,  $SD = 3.91$  months, age range: 73–83 months) were included in the final analyses, none of whom had participated in Experiment 1. The predominant socioeconomic background was middle class. Informed consent was obtained from the parents or teachers prior to their participation.

### 3.1.2. Procedure

In Experiment 2, participants were given the same tasks as in Experiment 1. The test was conducted in the same quiet room at the kindergarten in one session over a period of 45 minutes.

### 3.1.3 Measures

#### 3.1.3.1 Executive Functions

We conducted the same four tasks to measure executive functions as in Experiment 1.

#### 3.1.3.2 Toast-Making Task

In Experiment 2, the materials and procedures were nearly the same as in Experiment 1, with one exception: in addition to the ingredients used in Experiment 1 (toast, butter, and cheese), we added fried eggs, which were made of paper clay and placed in a white plastic box. Thus, the children were given a different toast recipe than in Experiment 1; the toast for the mouse contained butter, cheese, and fried eggs, because the mouse liked them, but the toast for the cat contained only butter and cheese because the cat liked butter and cheese, but not fried eggs. The

materials were lined up as follows: white box of toast, box of butter, box of cheese, box of fried eggs, and the toaster oven. Given this arrangement, when making toast for the mouse branch point errors could occur at the transition point from the subtask of adding cheese to the next subtask of adding fried eggs.<sup>4</sup>

We identified 28 trials as having imprecise time of interruptions, and these trials were excluded. The remaining trials were evaluated for the occurrence of branch point errors; we scored rates of branch point errors, including microslips, across 12 trials. All the trials from 59 children were coded by the experimenter and the same coder as for Experiment 1. The consistency rate was 92% ( $\kappa = .87$ ) between the two coders, and coding disagreements were resolved through discussion. The trials from other 13 children were coded by the experimenter alone.

## 3.2 Results and Discussion

### 3.2.1 Developmental Changes of Routines

In line with Experiment 1, we conducted the analysis with mixed-effects logistic regression using R statistical software. Preliminary analysis indicated no main effect of sex ( $b = -0.11, z = -0.98, \chi^2 = 0.95, df = 1, p = .330$ ); thus, the sex factor was excluded from subsequent analyses. As possible predictors, the following fixed-effect variables were included in the analysis: an age factor, an interruption factor (the comparison of the control condition with the average of the middle and end interruption conditions), a timing factor (the comparison of the

---

<sup>4</sup> In Experiment 2, we set a recipe for making toast which was less intuitive than that for Experiment 1 where a mouse liked cheese. To examine whether the difference in the recipes would have an impact on the performance of sequential actions, we conducted direct comparison between Experiment 1 and Experiment 2 only in 4- and 5-year-olds. Although the longer sequential actions and less intuitive recipe used in Experiment 2 were expected to cause children to produce more branch point errors than were those in Experiment 1, it was shown that there was no difference in performance in the control condition ( $F(1, 86) = 0.82, p = .367$ ). Thus, the less intuitive recipe was unlikely to affect children's performance in the toast-making task.



middle interruption condition with the end interruption condition), and two-way interactions. Only a random intercept for participant was included as a random effect variable, because the model that included the other random variables failed to converge. Additionally, the best-fitting model was determined based on AIC, and a likelihood ratio test was used for significance testing, as in Experiment 1.

The results of the best-fitting model ( $AIC = 749.5$ ), which fits significantly better than the null model ( $AIC = 772.5$ ,  $p < .001$ ), are shown in Table 4. Consistent with Experiment 1, the results of the analysis indicated significant main effects of the interruption factor and the timing factor (see Table 4). As in Experiment 1, the interruption led to more branch point errors; specifically, middle subtask interruption had a more disruptive impact on the performance of sequential actions than did end subtask interruption. Thus, it was suggested that overall, preschoolers exhibited degraded and recovered context representations over the course of a sequential action. However, we found no significant interaction between age and the timing factor ( $\chi^2 = 0.87$ ,  $df = 1$ ,  $p = .352$ ). In Experiment 2, as compared to Experiment 1, we could not find developmental differences in how preschoolers represented task contexts.

-----

Insert Table 4 about here

-----

### 3.2.2 Executive Functions and Developmental Changes of Routines

Our next aim was to validate the relationships between the timing factor, the interruption factor, and the advanced DCCS. To achieve this, we conducted the same mixed-effects logistic regression analysis as in Experiment 1. The following fixed-effect variables were included in the analysis: main effects of age; four executive function tasks; interruption factor; timing factor;

and the two-way interactions (advanced DCCS x interruption factor, advanced DCCS x timing factor). Random intercepts for participant and animal were included for random effect variables.

Table 5 shows the results of the best-fitting model (AIC = 739.5), which fits significantly better than the null model (AIC = 774.5,  $p < .001$ ). In accordance with Experiment 1, the model including the executive function tasks (Table 5) fit significantly better than the model exploring only age differences (Table 4) ( $\chi^2 = 15.99$ ,  $df = 3$ ,  $p = .001$ ). Importantly, we confirmed the interaction between the timing factor and the advanced DCCS (see Figure 3). Post-hoc analysis revealed that the error rate for the middle interruption condition was significantly higher than that in the end interruption condition in children with high performance on the advanced DCCS ( $b = 2.35$ ,  $z = 6.00$ ,  $p < .001$ ), and that the differences of interruption timing were also significant in children with low performance on the advanced DCCS ( $b = 0.85$ ,  $z = 2.38$ ,  $p = .017$ ).

Although both children with high and low performance on the advanced DCCS were more vulnerable to the middle subtask interruption, what was important that the significant interaction between the timing factor and the advanced DCCS indicated the differential influences of interruption timing between children who showed high and low performance in the advanced DCCS. Hence, even when the common subtask was added, the development of executive functions was again found to be associated with the differential modulation of context representations between the middle and the end of the subtask. Experiment 2, however, did not show a significant interaction between the interruption factor and the advanced DCCS ( $\chi^2 = 0.21$ ,  $df = 1$ ,  $p = .647$ ). This is possibly due to higher branch error rates in the middle interruption condition, which raised average branch point error rates irrespective of advanced DCCS performance.

These findings indicate that children with high performance on the advanced DCCS produced fewer branch point errors in the end interruption condition than did those with low performance on the advanced DCCS, whereas the error rate in the middle interruption condition was not strongly influenced by performance on the advanced DCCS, suggesting that executive functions measured by the advanced DCCS were associated with overcoming the end subtask interruption rather than being more vulnerable to the middle subtask interruption.

-----

Insert Table 5 about here

-----

-----

Insert Figure 3 about here

-----

## **5. General Discussion**

To date, many developmental studies have demonstrated that preschoolers exhibit dramatic increases over their developmental stage in their abilities to perfectly imitate and learn novel sequential actions (e.g., Buchsbaum, Gopnik, Griffiths, & Shafto, 2011; Flynn & Whiten, 2008; Lyons, Young, & Keil, 2007; Whiten, Flynn, Brown, & Lee, 2006). It has been also demonstrated that preschoolers can report what happens in familiar events and perform familiar sequential actions based on their own daily experiences (e.g., Freier et al., 2015; Hudson & Fivush, 1991; Yanaoka & Saito, 2017). However, it remains unclear what cognitive developmental mechanisms support learning to perform sequential actions as routines in young children.

Across two experiments, four- and five-year-olds quickly learned to produce novel sequential actions and were able to produce them routinely, as indicated by the low error rate and the non-significant effects of age in the control condition. These results are consistent with previous developmental studies showing that even three-year-olds perform familiar sequential actions (e.g., Freier et al., 2015; Hudson & Fivush, 1991; Yanaoka & Saito, 2017). Importantly, this is the first study demonstrating that while they performed a sequential action repeatedly, a representation of task contexts would change flexibly depending on its relevance to action selection. Notably, there were developmental differences in how young children represent task contexts and executive functions were related to these developmental changes. These findings fit with computational models for acquiring routines (e.g., Botvinick & Plaut, 2004; Cooper et al., 2014) and empirical evidence from adults (Botvinick & Blysm, 2005). In the following sections, we promote our understanding of how preschoolers learn and perform repeated sequential actions through the scrutiny of the results of Experiment 1 and Experiment 2.

### 5.1 Developmental Changes in Learning to Perform Routines

A key feature of our studies is that we tested the effect of interruptions at different times on selecting a context-dependent action in performing task sequences. Experiment 1 revealed that older children tended to be more vulnerable to middle subtask interruption than end subtask interruption, whereas the pattern among younger children was not as salient. Following Botvinick and Plaut (2004), this finding suggests that only older children can learn to perform routines in a similar manner to adults, that is, they represent the two task contexts as being more similar at the middle of the common subtask than at the end of the common subtask.

In contrast, Experiment 2, in which the sample was extended to six-year-olds, did not show the above developmental trends. One of the reasons for this finding could be the better

performance of the six-year-olds. As exploratory analyses, we separately conducted the mixed-effects logistic regression analysis for four- and five-year-olds and that for six-year-olds (see Figure A1). First, the analysis for four- and five-year-olds aimed to find the significant interaction between age and the timing factor as a replication of the finding from Experiment 1. The analysis was successful in replicating the significant interaction in the best-fitting model (see Table A1), which indicated that middle and end interruption differentially influenced on the performance of sequential action between in older children ( $b = 2.56, z = 4.92, p < .001$ ) and in younger children ( $b = 1.05, z = 2.20, p = .029$ ). The other analysis checked the main effects of the interruption and timing factors on performance only in the six-year-olds. It did not reveal a significant effect of the interruption factor (see Table A2); this indicated that our momentary interruptions were not demanding enough to interfere with the performance of sequential actions by six-year-olds. In fact, we also found a marginal significant main effect of the timing factor (see Table A2), and this effect might have been underestimated due to the small number of branch point errors. To directly test this possibility, future studies may examine whether giving six-year-olds a more demanding interruption (e.g., recalling supra-span digit lists three times) or a harder task (e.g., requiring them to make three different kinds of toast, for a cat, a mouse, and a dog) makes them vulnerable to middle subtask interruptions.

## 5.2 Executive Functions and Routines

Experiments 1 and 2 revealed that executive functions were closely related to the developmental changes in how preschoolers represented task contexts in executing task sequences. Children with high performance on the advanced DCCS, like adults, were more distracted by interruptions at the middle of the common subtask than at the end of the subtask, whereas children with low performance on the advanced DCCS did exhibit no such differences

(Experiment 1) or decreased differential influences (Experiment 2) by the two types of interruption. The goal circuit model developed by Cooper et al. (2014) may fit best with these findings. In this model, based on the neuropsychological evidence that different levels of hierarchical structure in behaviour are represented at different levels in a hierarchy of cortical areas (e.g., Badre, 2008; Bunge & Zelazo, 2006; Fuster, 2001, 2004; Koechlin, Ody, & Kouneiher, 2003), a goal unit was hierarchically added to a non-hierarchical simple recurrent network (Botvinick & Plaut, 2004). In addition, an executive control system may modulate an input of the current goal and feed back to the model recurrently; the recurrent connections allow the model to develop an implicit representation of task contexts through learning. Cooper et al. (2014) demonstrated that the goal circuit model could learn long, structurally complex sequential actions including a branch point more efficiently than the non-hierarchical simple recurrent network (Botvinick & Plaut, 2004), although the simple recurrent network could ultimately learn them, with more repetition. Thus, it is assumed that the development of executive functions (through goal-based learning) might facilitate the flexible modulation of task context representations during performance of sequential actions in our task.

Our results can also be interpreted from the viewpoint of hierarchical goal representations. The results from Experiment 2 suggested that executive functions were associated with overcoming the end subtask interruption rather than being more vulnerable to the middle subtask interruption. After children were interrupted at the end of the subtask, they returned to a desk and selected a context-dependent action for the next subtask. This required them to access a higher goal representation (i.e., making toast for the cat or the mouse), which should have been available at this time point. In this situation, success in selecting an appropriate context-dependent action depends on the success of the active maintenance of the higher goal

representation. Children's ability to maintain the higher goal representation actively is likely to be associated with executive functions (e.g., Munakata et al., 2012). In contrast, after being interrupted at the middle of the subtask, children first had to return to the interrupted subtask and therefore to access the sub-goal associated with that subtask in order to finish it. This access to the sub-goal representation at the middle of the subtask might impair later access to the higher goal representation, with the result that the higher goal representation would become more attenuated or degraded during the performance of the current subtask. This would have led to higher rates of branch point errors in all children. Probably the negative influence of access to the sub-goal on the status of the higher goal representation are not strongly related to children's executive functions. This account of goal representations is supported by the recent developmental evidence demonstrating the role of development in the maintenance of hierarchical goal representations and executive functions in performing sequential actions (Freier et al., 2015; Freier, Cooper, & Mareschal, 2017; Loucks & Melzoff, 2013; Loucks, Mutschler, & Meltzoff, 2017; Yanaoka & Saito, 2017). For example, Yanaoka and Saito (2017) reported that 5-year-olds developed the ability to maintain higher goal representation and showed more flexibility in the execution of scripts than 4-year-olds in a task where they helped a doll select items to wear. But if sub-goals were experimentally activated, both 5-year-olds and 4-year-olds reduced their flexible actions, based on the higher main goal. That is, access to sub-goals wiped out developmental and individual differences in executive functions that support main-goal directed behaviors.

Cooper et al. (2014) argued that the context representation account and the goal representation account were not mutually exclusive. At the beginning of the learning phase, their model was trained to attain both main goals and sub-goals, to bias the lower level system

towards task-appropriate actions. However, with practice, the information about the sequential actions must become re-represented as a distributed representation within the model's internal hidden layer, and the need for input from goal representations is reduced. In line with the model (Cooper et al., 2014), Ruh et al. (2010) argued that selecting the next action at a branch point required cognitive control, based on the fact that adults selected an action less quickly when they concurrently performed a secondary, attentionally demanding task than when they did not. Importantly, interference with the secondary task was reduced with increasing practice. As such, as sequential actions became more well-learned or routinized, they came to rely on context representations instead of goal representations. Our findings, however, did not provide strong evidence to support which representations supported children's sequential actions; still, our findings were consistent with Cooper et al. (2014), and we were able to extend the model to the developmental trajectory.

### 5.3 Methodological Implications

Our novel task for preschoolers will contribute to future research on routines. Previous developmental studies (e.g., Freier et al., 2015; Yanaoka & Saito, 2017) used materials and task contexts that children often experienced in their daily lives (e.g., making a sandwich, changing clothes for going to kindergarten). In such cases, their experience cannot be separated from an age factor; that is older children have more experience with the materials and task contexts than younger children. Thus, we used novel task contexts (e.g., making toast for a cat or a mouse according to original recipes) in our toast-making task. Although the materials used in the task were familiar to children, the task contexts were relatively independent of their daily experience. Consequently, the performance of the toast-making task reflects the preschoolers' ability to learn to perform sequential actions through the demonstration, practice trials, and test trials only in



experimental settings. One may notice that the design of this task is similar to pretend play; thus, there is a possibility that the ability to perform this task might be influenced by the ability to conduct pretend play. However, pretend play peaks at around 3–5 years with pretend identities (Lillard, Lerner, Hopkins, Dore, Smith, & Palmquist, 2013), and most of our sample were well able to engage in it. As such, the degree to which the children engaged in pretend play is unlikely to have had an impact on their task performance. Our original task, therefore, overcomes the limitation of the previous tasks used in developmental psychology and is appropriate for children.

#### 5.4 Conclusions

The current findings provide the first evidence in the literature of the developmental mechanisms underlying the acquisition of sequential actions in routine situations. We have demonstrated the developmental differences in how preschoolers represent task contexts during the execution of repeated task sequences. For instance, older children are able to modulate context representation more flexibly according to the task goal than younger children. Moreover, executive functions were found to be related to these developmental changes. We believe that our findings offer important insight into the relationships between explicit goal representation, context representation, and the acquisition of routines in preschoolers.

### **Acknowledgements**

The research reported here was supported by a Grant-in-Aid for Japan Society for the Promotion of Science (JSPS) fellows (16J0372). We are grateful to the staff and students of the kindergarten schools for their participation in this research.

### References

- Altmann, E. M., & Trafton, G. J. (2002). Memory for goals: an activation-based model. *Cognition*, *26*, 39-83. [https://doi.org/10.1016/S0364-0213\(01\)00058-1](https://doi.org/10.1016/S0364-0213(01)00058-1)
- Arnold, A., Wing, A. M., & Rotshtein, P. (2017). Building a Lego wall: Sequential action selection. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 847-852. <http://dx.doi.org/10.1037/xhp0000382>
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, England: Cambridge University Press.
- Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*, 193-200. <https://doi.org/10.1016/j.tics.2008.02.004>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9. <<https://CRAN.R-project.org/package=lme4>>.
- Bauer, P. J., & Hertzgaard, L. A. (1993). Increasing Steps in Recall of Events: Factors Facilitating Immediate and Long-Term Memory in 13.5-and 16.5-Month-Old Children. *Child Development*, *64*, 1204-1223. <https://doi.org/10.1111/j.1467-8624.1993.tb04196.x>

- Bauer, P. J., & Mandler, J. M. (1989). One thing follows another: Effects of temporal structure on 1-to 2-year-olds' recall of events. *Developmental Psychology, 25*, 197-206.  
<http://dx.doi.org/10.1037/0012-1649.25.2.197>
- 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.  
<http://dx.doi.org/10.1037/0012-1649.28.3.441>
- Bauer, P. J., & Thal, D. J. (1990). Scripts or scraps: Reconsidering the development of sequential understanding. *Journal of Experimental Child Psychology, 50*, 287-304.  
[https://doi.org/10.1016/0022-0965\(90\)90043-8](https://doi.org/10.1016/0022-0965(90)90043-8)
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*, 1641-1660. <https://doi.org/10.1111/j.1467-8624.2010.01499.x>
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*, 647-663. <https://doi.org/10.1111/j.1467-8624.2007.01019.x>
- Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology, 114*, 375-388. <https://doi.org/10.1016/j.jecp.2012.09.015>
- Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences, 12*, 201-208. <https://doi.org/10.1016/j.tics.2008.02.009>

- Botvinick, M. M., & Bylsma, L. M. (2005). Distraction and action slips in an everyday task: Evidence for a dynamic representation of task context. *Psychonomic Bulletin & Review*, *12*, 1011-1017. <https://doi.org/10.3758/BF03206436>
- Botvinick, M., & Plaut, D. C. (2002). Representing task context: proposals based on a connectionist model of action. *Psychological Research*, *66*, 298-311. <https://doi.org/10.1007/s00426-002-0103-8>
- Botvinick, M. M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, *111*, 395-429. <http://dx.doi.org/10.1037/0033-295X.111.2.395>
- Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*, 331-340. <https://doi.org/10.1016/j.cognition.2010.12.001>
- Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science*, *15*, 118-121. <https://doi.org/10.1111/j.0963-7214.2006.00419.x>
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032-1053. <https://doi.org/10.1111/1467-8624.00333>

- Chevalier, N., & Blaye, A. (2009). Setting goals to switch between tasks: Effect of cue transparency on children's cognitive flexibility. *Developmental Psychology, 45*, 782-797. <http://dx.doi.org/10.1037/a0015409>
- Cooper, R. P., Ruh, N., & Mareschal, D. (2014). The goal circuit model: A hierarchical multi route model of the acquisition and control of routine sequential action in humans. *Cognitive Science, 38*, 244-274. <https://doi.org/10.1111/cogs.12067>
- Cooper, R.P., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review, 113*, 887-916. <http://dx.doi.org/10.1037/0033-295X.113.4.887>
- Diamond, A., Prevor, M., Callender, G., & Druin, D. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU—introduction. *Monographs of the Society for Research in Child Development, 62*, 1-208. <http://dx.doi.org/10.2307/1166208>
- Fivush, R. (1984). Learning about school: the development of kindergartners' school scripts. *Child Development, 55*, 1697-1709. <http://dx.doi.org/10.2307/1129917>
- Freier, L., Cooper, R. P., & Mareschal, D. (2015). The planning and execution of natural sequence actions in preschool years. *Cognition, 144*, 58-66. <https://doi.org/10.1016/j.cognition.2015.07.005>
- Freier, L., Cooper, R. P., & Mareschal, D. (2017). Preschool children's control of action outcomes. *Developmental Science, 20*, e12354. <http://dx.doi.org/10.1111/desc.12354>

- Flynn, E., & Whiten, A. (2008). Imitation of hierarchical structure versus component details of complex actions by 3-and 5-year-olds. *Journal of Experimental Child Psychology, 101*, 228-240. <https://doi.org/10.1016/j.jecp.2008.05.009>
- Fuster, J. M. (2001). The prefrontal cortex—an update: time is of the essence. *Neuron, 30*, 319-333. [http://dx.doi.org/10.1016/S0896-6273\(01\)00285-9](http://dx.doi.org/10.1016/S0896-6273(01)00285-9)
- Fuster, J. M. (2004). Upper processing stages of the perception–action cycle. *Trends in Cognitive Science, 8*, 143-145. <http://dx.doi.org/10.1016/j.tics.2004.02.004>
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60. <http://dx.doi.org/10.1037/0033-2909.134.1.31>
- Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*, 493-498. <http://dx.doi.org/10.1111/2041-210X.12504>
- Hudson, J. A., & Fivush, R. (1991). Planning in the preschool years: The emergence of plans from general event knowledge. *Cognitive Development, 6*, 393-415. [https://doi.org/10.1016/0885-2014\(91\)90046-G](https://doi.org/10.1016/0885-2014(91)90046-G)
- Hudson, J. A., Shapiro, L. R., & Sosa, B. B. (1995). Planning in the real world: Preschool children's scripts and plans for familiar events. *Child Development, 66*, 984-998. <http://dx.doi.org/10.1111/j.1467-8624.1995.tb00917.x>

Humphreys, G. W., Forde, E. M. E., & Francis, D. (2000). The organization of sequential action.

In S. Monsell & J. Driver (Eds.), *Control of cognitive processes, attention and performance* (Vol. 18, pp. 427-442). Cambridge, MA: MIT Press.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*, 1181-1185.

<http://dx.doi.org/10.1126/science.1088545>

Lashley, K. (1951). The problem of serial order in behavior. In L. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-136). New York: Wiley.

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M.

(2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin*, *139*, 1-34. <http://dx.doi.org/10.1037/a0029321>

Loucks, J., & Meltzoff, A. N. (2013). Goals influence memory and imitation for dynamic human action in 36-month-old children. *Scandinavian Journal of Psychology*, *54*, 41-50.

<http://dx.doi.org/10.1111/sjop.12004>

Loucks, J., Mutschler, J., & Meltzoff, A. N. (2017). Children's representation and imitation of events: How goal organization influences 3-year-old children's memory for action sequences.

*Cognitive Science*, *41*, 1904-1933. <http://dx.doi.org/10.1111/cogs.12446>

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of

overimitation. *Proceedings of the National Academy of Sciences*, *104*, 19751-19756. <https://doi.org/10.1073/pnas.0704452104>



- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Henry Holt and Co.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology, 41*, 49-100.  
<https://doi.org/10.1006/cogp.1999.0734>
- Moriguchi, Y. (2014). The early development of executive function and its relation to social interaction: a brief review. *Frontiers in Psychology, 5*.  
<http://dx.doi.org/10.3389/fpsyg.2014.00388>
- Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control: three key transitions. *Current Directions in Psychological Science, 21*, 71-77.  
<https://doi.org/10.1177/0963721412436807>
- Nelson, K. (1986). *Event knowledge: Structure and function in development*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Reason, J. (1990). *Human Error*. Cambridge, UK: Cambridge University Press.
- Ruh, N., Cooper, R. P., & Mareschal, D. (2010). Action selection in complex routinized behaviors. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 955-975. <http://dx.doi.org/10.1037/a0017608>

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ:

Lawrence Erlbaum.

Schwartz, M. F., Reed, E. S., Montgomery, M., Palmer, C., & Mayer, N. H. (1991). The quantitative description of action disorganization after brain damage: A case study. *Cognitive Neuropsychology*, 8, 381-414. <http://dx.doi.org/10.1080/02643299108253379>

Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society of London*, 351, 1405-1411. <https://doi.org/10.1098/rstb.1996.0124>

Shapiro, L. R., & Hudson, J. A. (2004). Effects of internal and external support on preschool children's event planning. *Journal of Applied Developmental Psychology*, 25, 49-73. <https://doi.org/10.1016/j.appdev.2003.11.004>

Simpson, A., & Riggs, K. J. (2005). Inhibitory and working memory demands of the day-night task in children. *British Journal of Developmental Psychology*, 10, 1-17. <https://doi.org/10.1348/026151005X28712>

Whiten, A., Flynn, E., Brown, K., & Lee, T. (2006). Imitation of hierarchical action structure by young children. *Developmental Science*, 9, 574-582. <https://doi.org/10.1111/j.1467-7687.2006.00535.x>

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, 108, 436-452. <https://doi.org/10.1016/j.jecp.2010.08.008>

- Yanaoka, K. (2014). Effects of planning and executive functions on turning back by young children in the execution of scripts. *Japanese Journal of Developmental psychology*, 25, 232-241. (In Japanese) <http://doi.org/10.11201/jjdp.25.232>
- Yanaoka, K., & Saito, S. (2017). Developing control over the execution of scripts: The role of maintained hierarchical goal representations. *Journal of Experimental Child Psychology*, 163, 87-106. <https://doi.org/10.1016/j.jecp.2017.06.008>
- Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, 1, 297-301. <https://doi.org/10.1038/nprot.2006.46>
- Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., ... & Carlson, S. M. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, i-151. Retrieved from <http://www.jstor.org/stable/1166202>

Table 1

*Experiment 1: Results of Mixed-Effects Logistic Regression in the Best-fitting Model for Age Differences*

	Estimate	SE	<i>z</i>	$\chi^2$	Pr ( $>\chi^2$ )
Intercept	1.720	.25	6.88		
Age	0.001	.02	0.07	0.005	.941
Timing	0.626	.20	3.10	9.58	.002
Interruption	0.836	.23	3.63	13.58	<.001
Timing $\times$ Age	0.047	.03	1.80	3.22	.073

Table 2

*Correlation Coefficients Among Four Executive Functions Tasks and their Correlation with Correct Rates of the Toast-Making Task in Each Condition*

	Red/blue task	Standard DCCS	Advanced DCCS	Correct rate (Control)	Correct rate (Middle)	Correct rate (End)
Red/blue task	-	-	-	.19	.27	.31
Standard DCCS	.25	-	-	-.20	-.10	-.01
Advanced DCCS	.37*	.31	-	-.05	.22	.45**
Nine box task	.11	.18	.32*	.21	.02	.24

\* $p < .05$ , \*\* $p < .01$

Table 3

*Experiment 1: Results of Mixed-Effects Logistic Regression in the Best-fitting Model for Effects of Executive Functions*

	Estimate	SE	<i>z</i>	$\chi^2$	Pr ( $>\chi^2$ )
Intercept	1.894	.26	7.40		
Timing	0.734	.22	3.41	12.11	<.001
Interruption	0.809	.23	3.52	13.67	<.001
Red-blue task	0.106	.05	2.06	4.05	.044
Standard DCCS	-0.278	.14	-1.98	3.79	.052
Advanced DCCS	0.130	.07	1.80	3.05	.081
Timing $\times$ Advanced DCCS	0.288	.13	2.14	4.88	.028
Interruption $\times$ Advanced DCCS	-0.264	.12	-2.21	4.65	.031

Table 4

*Experiment 2: Results of Mixed-Effects Regression in the Best-fitting Model for Age Differences*

	Estimate	SE	<i>z</i>	$\chi^2$	Pr ( $>\chi^2$ )
Intercept	1.613	.14	11.44		
Age	0.020	.01	1.54	2.31	.128
Timing	1.503	.29	5.12	26.28	<.001
Interruption	0.849	.33	2.60	7.05	.008

Table 5

*Experiment 2: Results of Mixed-Effects Logistic Regression in the Best-fitting Model for the Effect of Executive Functions*

	Estimate	SE	<i>z</i>	$\chi^2$	Pr ( $>\chi^2$ )
Intercept	1.672	.14	11.68	<.001	
Timing	1.596	.30	5.33	28.83	<.001
Interruption	0.768	.33	2.31	5.53	.019
Red-blue task	0.102	.06	1.74	2.91	.086
Advanced DCCS	0.107	.07	1.48	2.10	.148
Timing $\times$ Advanced DCCS	0.426	.13	3.37	11.65	<.001



## Figure Captions

*Figure 1.*

Materials for the toast-making task used in Experiment 1. Toast, butter, and cheese are kept in each white box. The object at the right end is a toaster oven.

*Figure 2.*

Mean percentages of branch point errors for the control, the middle interruption, and the end interruption condition in children with high and low performance on the advanced DCCS who took part in Experiment 1.

*Figure 3.*

Mean percentages of branch point errors for the control, middle interruption, and end interruption conditions in children with high and low performance on advanced DCCS who took part in Experiment 2.

Figure 1



Figure 2

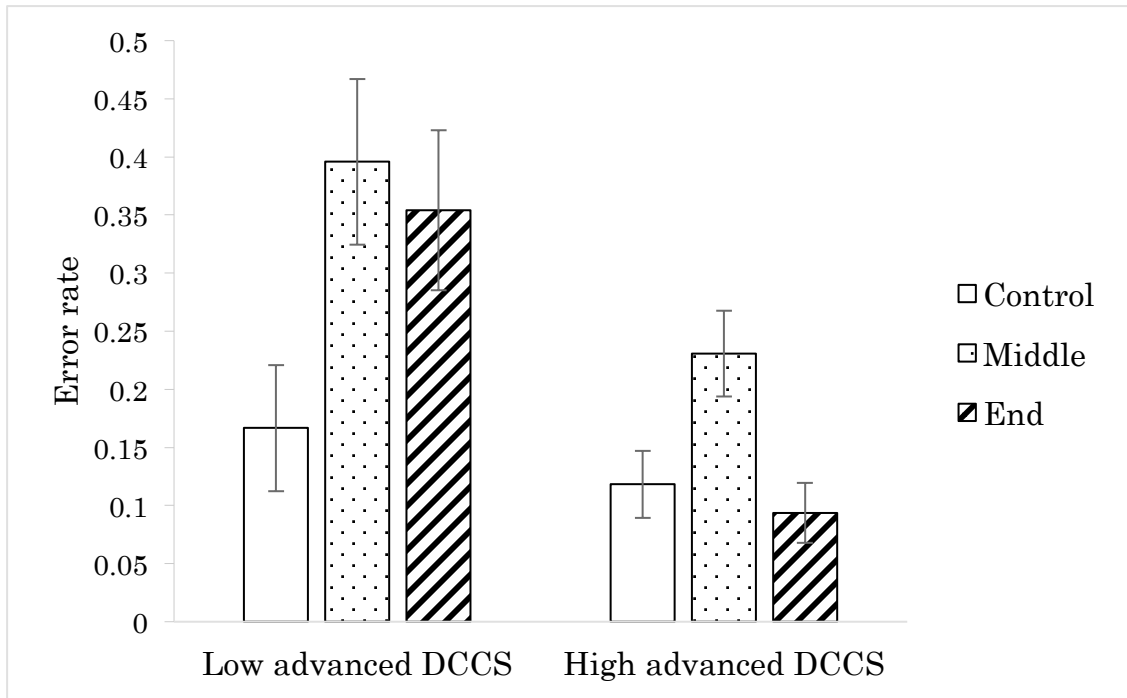
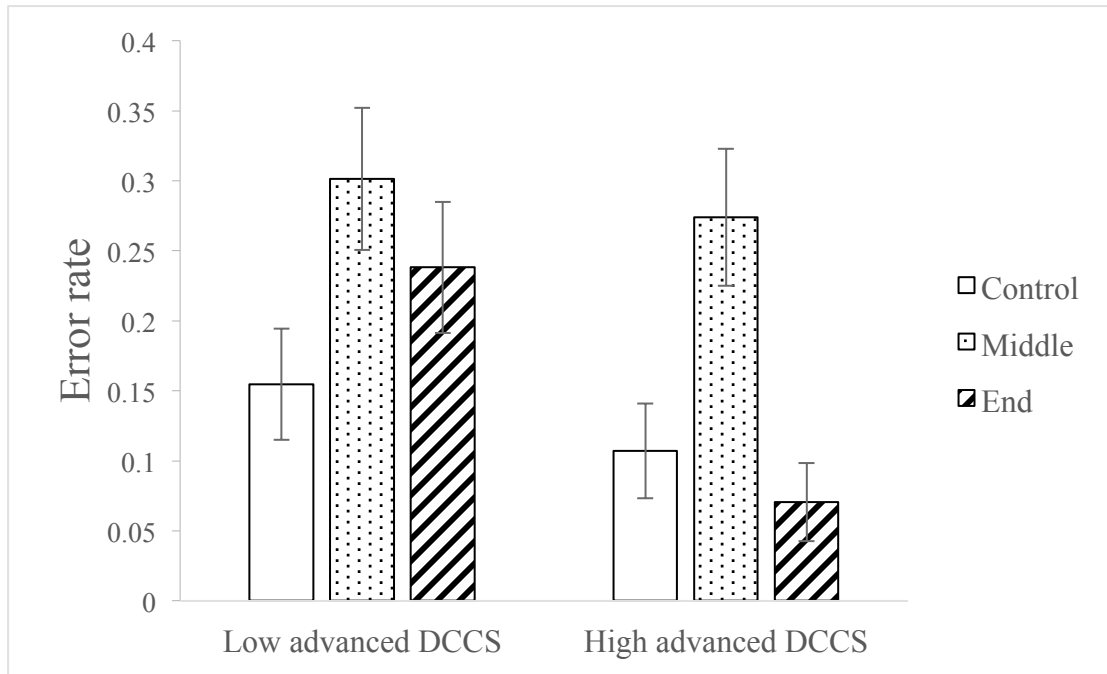


Figure 3



## Appendix

## Results of exploratory analyses for Experiment 2

Table A1

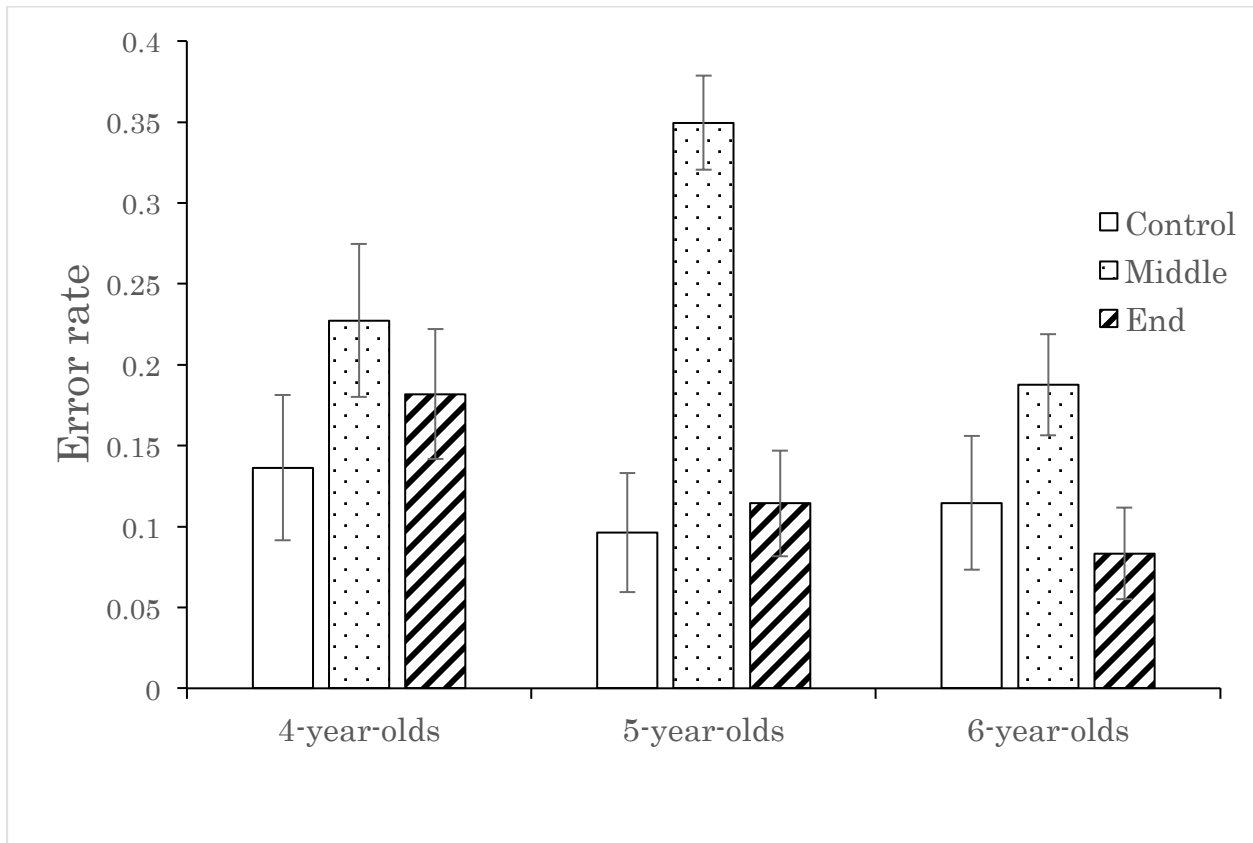
*Experiment 2: Results of Logistic Mixed-Effects Regression in the Best-fitting Model for Four- and Five-Year-Olds*

	Estimate	SE	z	$\chi^2$	Pr ( $>\chi^2$ )
Intercept	1.455	.17	8.55	<.001	
Age	-0.009	.03	-0.35	0.12	.726
Timing	1.793	.36	4.96	25.28	<.001
Interruption	1.166	.41	2.84	8.75	.003
Timing $\times$ Age	0.153	.06	2.74	7.64	.006
Interruption $\times$ Age	0.099	.06	1.57	2.49	.115

Table A2

*Experiment 2: Results of Logistic Mixed-Effects Regression in the Model for Six-Year-Olds*

	Estimate	SE	z	$\chi^2$	Pr ( $>\chi^2$ )
Intercept	1.99	.26	7.61		
Age	0.023	.06	0.41	0.16	.686
Timing	0.993	.54	1.86	3.37	.066
Interruption	0.270	.58	0.47	0.21	.641



*Figure A1.* Mean percentages of branch point errors for the control, middle interruption, and end interruption conditions in each age group.