Journal of Experimental Child Psychology 125 (2014) 85–101

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

# Development of planning in 4- to 10-year-old children: Reducing inhibitory demands does not improve performance

Emma C. Tecwyn *, Susannah K.S. Thorpe [1], Jackie Chappell [1]

*School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*

ARTICLE INFO

*Article history:*
Received 3 October 2013
Revised 19 February 2014

*Keywords:*
Planning
Inhibitory control
Problem solving
Cognitive development
Cognitive strategies
Physical cognition

ABSTRACT

Currently, there are relatively few tasks suitable for testing planned problem solving in children. We presented 4- to 10-year-old children ($N = 172$) with two planning tasks (sequential planning and advance planning) using the paddle-box apparatus, which was originally designed to investigate the planning skills of nonhuman apes. First, we were interested in the development of children's performance in the two tasks and whether the strategies children used to succeed differed among age groups. Performance improved significantly across age groups in both tasks. Strategies for success in the advance planning task differed among age groups, with 4- and 5-year-olds performing more excess actions, and a greater proportion of irrelevant excess actions, than older children. Findings are discussed in relation to the development of performance in tower tasks, which are a commonly used test of planning ability in humans. Second, based on previous findings with apes, we predicted that introducing measures to reduce the inhibitory demands of the advance planning task would improve children's performance. Therefore, in this study we introduced two methodological alterations that have been shown to improve children's performance in other tasks with inhibitory demands: (a) imposing a short delay before a child is allowed to act and (b) replacing reward items with symbolic tokens. Surprisingly,

---

* Corresponding author. Current address: Centre for Social Learning & Cognitive Evolution, School of Psychology and Neuroscience, University of St. Andrews, St. Andrews, Fife KY16 9JP, UK.

*E-mail address:* ect4@st-andrews.ac.uk (E.C. Tecwyn).

[1] Joint senior authors.

http://dx.doi.org/10.1016/j.jecp.2014.02.006

neither of these measures improved the performance of children in any of the age groups, suggesting that, contrary to our prediction, inhibitory control might not be a key performance-limiting factor in the advance planning paddle-box task.

---

## Introduction

Planned behavior involves considering different sequences of action alternatives and choosing between them prior to acting (McCormack & Atance, 2011). It is a complex cognitive process, the development of which occurs in conjunction with and is supported by various key executive function processes. Executive functions are a group of skills necessary for the control of thought and action required for reasoning, planning, and problem solving (Anderson, 1998; Baughman & Cooper, 2007). Key executive functions include inhibitory control, working memory, and task switching (Asato, Sweeney, & Luna, 2006; Baughman & Cooper, 2007; McCormack & Atance, 2011). Given that there are currently few appropriate tasks for investigating planned problem solving in young children (McCormack & Atance, 2011), this study used a novel paradigm originally designed for testing nonhuman apes (hereafter apes) to present two planning tasks to children and explored whether inhibitory control might be a key performance-limiting factor.

### *Development of planned problem solving*

Tower tasks, such as the Tower of London (ToL; Shallice, 1982), are the most commonly used test of planning ability in humans (McCormack & Atance, 2011), although other experimental studies of planning in children have used route planning tasks (e.g., Gauvain & Rogoff, 1989) and "real-world" scenarios (e.g., Hudson & Fivush, 1991). In the ToL task, participants are presented with two sets of three pegs (start and goal), with three different-colored discs arranged on each set. The aim of each trial is to rearrange the discs on the start pegs so that they match the configuration of discs on the goal pegs. Problem complexity can be manipulated by increasing the number of moves required to solve the trial as well as by altering structural features of the problem such as the number of intermediate moves (i.e., where a disc must be moved into a temporary position that is not its final goal position) a participant is required to make (Kaller, Rahm, Köstering, & Unterrainer, 2011). The planning demands of the task stem from the need to anticipate the consequences of one's next actions (McCormack & Atance, 2011). Efficient performance involves mental representation of the path from the start state to the goal state, followed by behavioral reproduction of the action sequence (Albert & Steinberg, 2011).

Several studies have examined the development of ToL performance across different age groups. Luciana and Nelson (1998) found that 4-year-old children performed poorly in 3-step ToL trials compared with 2-step trials. Asato and coworkers (2006) investigated the development of performance in 2- to 5-step ToL problems in 8- to 30-year-olds. Age effects were found only for the more complex 4- and 5-step problems. As age increased, the number of excess moves being made decreased (Asato et al., 2006). Finally, in a study of 10- to 30-year-olds, Albert and Steinberg (2011) found that performance in 3-step ToL trials was not mature until 16 or 17 years of age.

Work by Kaller and colleagues (Kaller, Rahm, Spreer, Mader, & Unterrainer, 2008; Kaller et al., 2011) demonstrates the significant impact that structural details of seemingly comparable ToL trials (rather than just the number of steps) can have on performance. For example, Kaller and colleagues (2008) showed that although 4-year-olds were able to solve 3-step trials that did not require an intermediate move, they struggled when an intermediate move formed part of the solution despite being instructed to plan their moves before starting. The authors suggested that the key difference between these two types of 3-step trial was that whereas trials without intermediate moves could be solved using a step-by-step perceptually guided strategy (i.e., where perceptual information can be used to guide behavior one step at a time), those requiring an intermediate move necessitated planning in terms of searching ahead to anticipate the consequences of at least the first two steps. The authors further suggested that development of performance may have been related to a switch in strategy, from a perceptually guided one to planning ahead (Kaller et al., 2008).

Children under 5 years of age find it particularly difficult to follow task instructions in tower tasks (Baughman & Cooper, 2007) or to operate commonly used computerized interfaces such as those used for maze navigation planning tasks (Miyata, Itakura, & Fujita, 2009). There is, therefore, a need to develop tasks that could be used to test the planning abilities of young children and infants. Furthermore, there is a lack of tasks that could potentially be used for cross-species comparisons, which is important if we want to understand something about the evolution of multi-step problem solving and begin to elucidate which underlying cognitive mechanisms, if any, humans share with other animals.

### *The paddle-box paradigm*

Tecwyn, Thorpe, and Chappell (2013) developed a novel paradigm (the paddle-box) to investigate the planning abilities of apes. The paddle-box consists of a transparent Perspex box containing eight rotatable paddles on three levels. At the bottom of the paddle-box are four possible goal locations that can each be configured as either open or blocked. Captive orangutans and bonobos were presented with two different tasks: sequential planning and advance planning. In both tasks, the aim of each trial was to turn a number of paddles to move a food reward from its starting position on one of the paddles inside the apparatus (the start paddle) to the open goal at the bottom. The reward could be successfully retrieved in a minimum of one, two, or three paddle rotations (classified as 1-, 2-, or 3-step trials). The key difference between the tasks was that for 2- and 3-step trials in the sequential planning task participants could always rotate the start paddle immediately and still potentially succeed. In the advance planning task, on the other hand, participants needed to pre-position one or two other paddles *before* rotating the start paddle because if the start paddle was rotated immediately the reward would become trapped. Both ape species performed well in the sequential planning task but generally failed in the advance planning task because they did not pre-position the relevant paddles before turning the start paddle. Tecwyn and colleagues (2013) suggested that the apes' poor performance in the advance planning paddle-box task may have been due to their difficulty in inhibiting the prepotent response of turning on the paddle with the highly salient food reward immediately and, hence, failing to perform the appropriate response of pre-positioning the paddles necessary for success (Tecwyn et al., 2013).

### *A role for inhibitory control in planning tasks?*

It is often the case that a task designed to test for a particular cognitive ability (e.g., planning) simultaneously taxes other mechanisms (Seed, Seddon, Greene, & Call, 2012). Inhibitory control is one such mechanism and is the ability to stop an inappropriate prepotent response or to ignore irrelevant information (Simpson & Riggs, 2007). An example of a prepotent response is reaching directly for a desirable object (Diamond, 1990). Responses may be prepotent because they are biologically predisposed, they are afforded by particular objects, they are habitual, or they are associated with desirable consequences (Simpson & Riggs, 2007). Inhibitory abilities develop slowly and are not fully mature until early adulthood (Diamond, 2002).

There is evidence that inhibitory control affects performance in tower tasks. Although the ToL does not establish a strong prepotent response that needs to be inhibited in terms of a salient reward item (unlike in the paddle-box task), participants need to inhibit making tempting "trap" moves (e.g., placing a disc in its goal position when an intermediate move is required) and delay immediate impulsive responding in favor of planning (Albert & Steinberg, 2011). In Kaller and colleagues' (2008) study, 4-year-olds' difficulty in solving 3-step trials requiring an intermediate move may have been related to their inability to inhibit making an impulsive but inappropriate first step of placing a disc onto its goal peg. Asato and colleagues (2006) found that increased success in 4- and 5-step ToL trials was significantly associated with fewer errors in an eye movement test of response inhibition in 8- to 13-year-olds and attributed this to the continuing development of voluntary cognitive control (e.g., the ability to switch between tasks or ignore distracting information) into adolescence. In Albert and Steinberg's (2011) study of ToL performance in 10- to 30-year-olds, the authors found that impulse control was the best predictor of performance.

### *Reducing the inhibitory demands of tasks: Delays and tokens*

Certain methodological alterations have been shown to reduce the inhibitory demands of tasks and, hence, improve performance. In this study, we tested the effectiveness of two particular measures that it is possible to implement with the paddle-box apparatus: imposing a short delay before permitting a child to respond, and replacing rewards with symbolic tokens.

A well-known test of inhibitory control is the *day–night task* (Gerstadt, Hong, & Diamond, 1994), in which children are required to say "day" to a picture of a moon and to say "night" to a picture of a sun. Both 3- and 4-year-olds perform poorly in this task because they struggle to inhibit saying what the stimuli really represent (Gerstadt et al., 1994). Diamond, Kirkham, and Amso (2002) found that implementing a delay before children were allowed to respond—during which time the experimenter sang a short rhyme—enabled 3- and 4-year-olds to succeed. The *go/no-go task* (Livesey & Morgan, 1991) has also been employed to investigate the development of inhibitory control in children. In this paradigm, children are presented with a series of boxes and told that boxes with a particular cue on the lid contain a reward and should be opened ("go" trials), whereas boxes with a different cue on the lid are empty and should be left shut ("no-go" trials). Although 3- and 4-year-olds succeed in opening boxes on go trials, they also frequently incorrectly open boxes on no-go trials even when instructions are made explicit and there are negative consequences for doing so (Simpson & Riggs, 2007). In a version of the go/no-go task, Simpson and colleagues (2012) found that introducing a 2-s delay between presenting the box and placing the cue on the lid significantly improved children's performance on no-go trials. The authors argued that increasing the time between presenting the triggering stimulus (the box) and allowing a child to respond permits passive dissipation of the prepotent response to open the box, enabling formation of an appropriate response strategy (Simpson et al., 2012). Mitchell and Potson (2001) compared the performance of two groups of adults in a set of 5- and 6-step ToL trials, all of which required inhibition of one "tempting" but inappropriate move. Participants in the experimental group were told to stop and think about certain moves, whereas participants in the control group were not. Inducing this delay significantly improved performance in terms of the number of trials completed in the minimum number of steps (Mitchell & Potson, 2001).

The *windows task* (Russell, Mauthner, Sharpe, & Tidswell, 1991) has similar behavioral inhibition demands to the go/no-go task. Here, each child plays with an opponent and is presented with two boxes with transparent lids: one that contains a visible reward and one that is empty. The child is told to point to the box that he or she wants the opponent to get, and the child then gets to open the other box himself or herself. In this task, 3-year-old children typically and repeatedly fail (they point at the box containing the reward and, therefore, lose it to their opponent), whereas 4-year-olds succeed. Hala and Russell (2001) suggested that 3-year-olds' difficulty in this task is related to their needing to inhibit the prepotent response of pointing at the desirable reward while holding the rules of the task in mind. Apperly and Carroll (2009) found that replacing the reward (stickers) in the windows task with any one of five types of symbol (token), including equally desirable sweets and a photograph of stickers, significantly improved the performance of 3- and 4-year-olds. The authors concluded that because the children's decision making is dominated by the desire to obtain the reward, symbols enable them to avoid impulsive responding and think more flexibly, permitting an alternative appropriate response to be formulated (Apperly & Carroll, 2009).

The aims of this study were twofold. First, we wanted to examine whether the paddle-box paradigm would detect age differences in performance in the sequential planning task and/or advance planning task (in terms of both success and strategy used) and to interpret our findings in the context of the existing literature on the development of performance in tower tasks given that they are well-established tests of planning ability. If results are comparable, this would suggest that the paddle-box is an appropriate paradigm for investigating the development of planning in children. We predicted that even the youngest children (4- and 5-year-olds) would perform well in the sequential planning paddle-box task given that 4-year-olds succeed in simple 3-step ToL trials (Kaller et al., 2008). Furthermore, orangutans and bonobos were capable of success in sequential planning trials (Tecwyn et al., 2013), and recent work by Völter and Call (2014) showed that apes performed comparably to 4- and 5-year-old children in a vertical maze planning task. We expected performance in the advance planning task (particularly in 2- and 3-step trials) to improve with increasing age, but possibly to not reach ceiling given that planning ability continues to develop into adolescence (Asato et al., 2006) and 3-step ToL performance might not be mature until 16 or 17 years of age (Albert & Steinberg, 2011). We were also interested in whether latency to first move would vary (a) across age groups, which has been reported for 4- and 5-step ToL trials (Albert & Steinberg, 2011; Asato et al., 2006; Luciana et al., 2009), and (b) with increasing minimum number of moves (i.e., task complexity), which is also in keeping with findings from ToL studies (Albert & Steinberg, 2011; Luciana et al., 2009).

Second, we investigated whether inhibitory control was a key factor limiting 4- to 10-year-olds' performance in the advance planning paddle-box task by introducing two methodological alterations that have been shown to improve performance in other tasks with inhibitory demands: (a) implementing a short delay (2 s, based on the findings of Simpson et al., 2012, in the go/no-go task) before children were allowed to respond (delay condition) and (b) replacing stickers with symbolic tokens (tokens condition). We predicted that if inhibitory control was key to successful performance in the advance planning task (because the prepotent desire to rotate the start paddle immediately needs to be inhibited), more children would succeed in the delay and tokens conditions (where inhibitory demands are reduced) compared with in a control condition. Specifically, we proposed that fewer children should rotate the start paddle immediately and instead succeed in pre-positioning relevant paddles in the experimental conditions. We predicted that the experimental conditions should have no effect on performance in the sequential planning task because in this task children could always rotate the start paddle straight away and still potentially succeed.

## Method

### Participants

The final sample consisted of 172 children: 60 4- and 5-year-olds (30 boys and 30 girls, mean age = 5 years 0 months [5;0], range = 4;6–5;6), 60 6- and 7-year-olds (31 boys and 29 girls, mean age = 7;1, range = 6;7–7;6), and 52 9- and 10-year-olds (26 boys and 26 girls, mean age = 10;0, range = 9;7–10;6) from three primary schools in the Birmingham area of the United Kingdom. An additional two children (both boys, one from the 4–5-year age group and one from the 6–7-year age group) started the experiment but dropped out partway through due to a lack of motivation to participate; their data were excluded from analyses. The ethnic composition of the final sample was 72% Asian, 15% Caucasian, 7% Black, and 6% other/unknown.

### Materials

The paddle-box apparatus (Tecwyn et al., 2013), originally designed to investigate the planning abilities of apes, was used. The paddle-box consisted of an opaque Perspex box (60 × 60 × 6 cm) containing eight rotatable paddles (14.5 × 3.5 × 1.7 cm; labeled 1–8 in Fig. 1) on three levels. There were four possible goal locations (each measuring 11 × 4.5 × 4.5 cm; labeled A–D in Fig. 1) at the base of the apparatus that could be either open or blocked. On each trial, children needed to try to get an item (a piece of sponge with a sticker on it or just a piece of sponge [2 × 2 × 2 cm], depending on the experimental condition) from the paddle on which it was placed by the experimenter (start paddle) to the open goal location (on each trial, there was only one open goal location). Children rotated the paddles using wooden handles (7 × 2.5 × 1.7 cm) that extended out of the front of the box and were oriented parallel to the paddles inside the box (see Fig. 1). The handles could be operated in a number of ways, for example, by pushing down from above or up from underneath at either end of a handle or by using

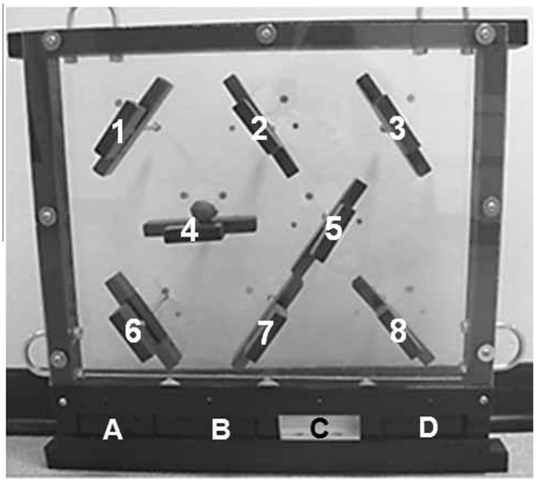

**Fig. 1.** Photograph of the paddle-box apparatus from children's viewpoint showing the eight rotatable paddles (labeled 1–8) on three levels and four possible goal locations (labeled A–D) at the bottom. The paddles could be held in one of three positions by weak magnets: flat (Paddle 4), diagonal left (e.g., Paddle 1), or diagonal right (e.g., Paddle 2). The apparatus as depicted is configured for an advance planning trial; the item to be retrieved is located on Paddle 4 (start paddle) and the open goal location is in the C position, whereas the other three goals (A, B, and D) are blocked.

a twisting action. They were designed to be large enough so that they did not require fine motor control and, thus, reduced the chance of children accidentally turning them the wrong way. Once a paddle was rotated, directional choices were not easily correctable because the item slid off the paddle. Paddles were held in position by weak magnets (Fig. 1) so that they were easily rotatable by children, but a moving reward did not displace them from their orientation.

### *Design*

A mixed design was used, with task (advance or sequential planning) and minimum number of steps (1, 2, or 3) as within-participant factors and condition (control, delay, or tokens) and age group as between-participant factors. The dependent variables were number of trials correct in each task; percentage of participants succeeding in at least one 1-, 2-, or 3-step trial; percentage of participants only rotating the start paddle; number of excess paddle rotations; and latency to first move.

### *Procedure*

Children were tested in a quiet area close to their classroom. The experiment was presented as a game where stickers could be won. Children sat at a table with the paddle-box and two experimenters. The main experimenter ran the experiment and coded the data, and the second experimenter assisted with configuring the apparatus between trials. Children were alternately assigned to one of three conditions: control, delay, or tokens. First, children completed the warm-up exercise to ensure that they were able to operate the simple paddle mechanism and retrieve the item from an open goal. All children were then presented with six trials of two different planning tasks: sequential planning and advance planning (12 trials in total). The order in which the two tasks were presented to children was counterbalanced. Within each task, children were presented with the same set of six trials in the same order. Each trial was a unique configuration of the apparatus, with the minimum number of steps required to retrieve the item (1, 2, or 3), start level, and position of the goal location pseudorandomized and the constraint that none of them was the same for more than two consecutive trials. For each task, there were two 1-step, two 2-step, and two 3-step trials.

### *Warm-up task*

Children were presented with the apparatus with the B and C goals open and the A and D goals blocked. They were then given one of two sets of instructions, depending on the condition to which they were assigned:

> *Control and delay conditions*—"Every time you manage to get a sticker out of the puzzle, you can keep it, okay? You can touch anything on this side of the puzzle [*experimenter indicates front of the paddle-box*]. Let's have a practice first."
> *Tokens condition*—"Every time you manage to get a sponge out of the puzzle, you will win a sticker, okay? You can touch anything on this side of the puzzle [*experimenter indicates front of the paddle-box*]. Let's have a practice first."

The sticker/sponge was then placed on Paddle 7 (Fig. 1, bottom center), and children were asked to try and get it out of the apparatus. If children failed to do this spontaneously, they were given neutral prompts such as "Can you think how you might be able to get the sticker/sponge out?" If they still failed, it was demonstrated to them by the experimenter, and the warm-up task was repeated until children spontaneously retrieved the sticker/sponge.

### *Testing phase*

Participants in all three conditions were given the following instructions:

> "Okay, now that we've had a practice, we're going to play the game lots more times so you can try to win some more stickers! I'm going to ask you to cover your eyes and not look each time I set up a new game so it's a surprise. Is that okay?"
>
> The next set of instructions varied between conditions as follows:
>
> *Control condition*—"When I tell you to open your eyes, you can try to get the sticker."
> *Delay condition*—"When I tell you to open your eyes, you can look at the puzzle, but please wait until I say 'go' before you try to get the sticker" [*experimenter implements a 2-s delay*].
> *Tokens condition*—"When I tell you to open your eyes, you can try to get the sponge."

Having configured the apparatus, the sticker/sponge was placed on the start paddle by the experimenter. Stickers were oriented so that they were facing the children. If a sticker was retrieved in the control or delay condition, children peeled the sticker off the sponge and kept it. If a sponge was retrieved in the tokens condition, children passed it to the experimenter in exchange for a sticker selected by the experimenter. If the item became trapped at one of the blocked goal locations in any of the conditions, it was removed from the back of the apparatus by the experimenter and children did not receive a sticker for that trial.

### *Sequential planning task*

In the sequential planning task, all of the paddles were set up in a flat orientation at the start of each trial. The number of steps required to solve each trial was dictated by the level on which the reward started. Each trial could be solved in a minimum of one, two, or three steps. All trials could be solved by rotating the start paddle first and then by rotating paddles on which the reward was subsequently located (see Fig. 2A for an example of a 2-step sequential planning trial).

### *Advance planning task*

In each trial, only the start paddle was positioned in a flat orientation; all of the other paddles were positioned in one of the two possible diagonal orientations (Figs. 1 and 2B). This meant that in 2- and 3-step trials, if the start paddle was rotated first, the reward would slide down to the bottom of the

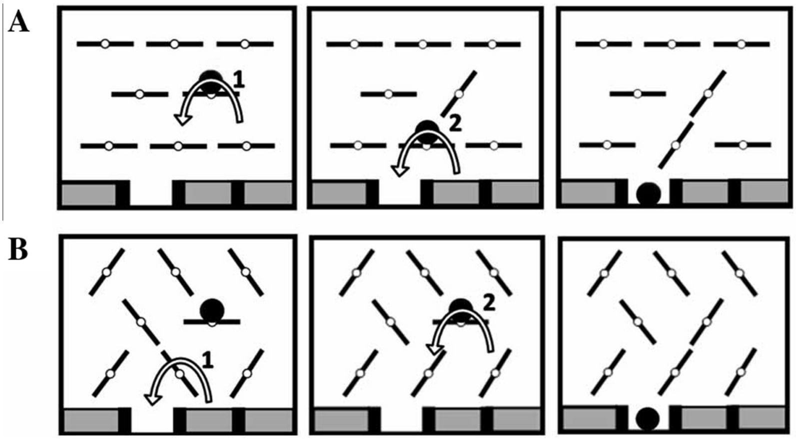

**Fig. 2.** Schematic examples of how to solve a 2-step sequential planning trial (A) and a 2-step advance planning trial (B) in the minimum number of moves.

apparatus and end up at one of the blocked goal locations. Therefore, the key difference between this task and the sequential planning task was that here, in 2- and 3-step trials, children needed to pre-position one or two other paddles *before* rotating the start paddle (Fig. 2B). In 1-step trials, it was not necessary to pre-position any other paddles, but the reward could start on any of the three levels. As in the sequential planning task, each trial could be solved in a minimum of one, two, or three steps.

## *Coding and analysis*

Data were scored live. For each trial, whether the sticker/sponge was successfully retrieved from the open goal or became trapped at one of the blocked goal locations was recorded. The measure of overall performance was the number of correct trials out of six in each task. To examine performance differences between trial types (1-, 2-, and 3-step), the dependent measure was the percentage of children succeeding in at least one of two trials of that type. To gain a better understanding of why children might be failing in the advance planning task, the number of children who only rotated the start paddle in every one of their trials was also recorded. Strategies used in the advance planning task were examined by scoring the number of excess paddle rotations performed (i.e., those that exceeded the minimum number of rotations in which the reward could be retrieved) together with whether each of these paddle rotations was relevant or irrelevant for retrieving the sticker/token. Relevant paddles were defined as those that could affect the path between the start paddle and the open goal (e.g., Paddle 7 in Fig. 1), whereas irrelevant paddles could not (e.g., Paddle 6 in Fig. 1). Latency to first move was recorded using a stopwatch and was defined as the time between children opening their eyes and children performing their first paddle rotation. In the delay condition, latency to first move included the 2-s experimenter-imposed delay.

Data were analyzed using IBM SPSS Statistics 21. Preparation time data were log-transformed to enable parametric statistics to be used. When it was not possible to use parametric statistics because transformation did not result in normally distributed data (as was the case for number of trials correct in each task and number of excess paddle rotations), nonparametric tests were used. Chi-square tests were used to look for differences among the three age groups in terms of the percentage of children succeeding in at least one 1-, 2-, and 3-step trial in the two tasks as well as the percentage of children only rotating the start paddle in the advance planning task. In cases where more than 20% of cells in contingency tables had expected frequencies of less than 5, a Fisher's exact test was used instead of a chi-square test because it is appropriate for examining the significance of associations regardless of sample characteristics. All statistical tests were two-tailed, and the significance level of alpha was .05.

## Results

Two key questions were investigated. First, we were interested in whether the paddle-box paradigm would detect age-related trends in performance in the sequential planning and/or advance planning tasks. Specifically, we examined whether older children succeeded in more trials than younger children in both tasks and whether strategies used to succeed in the advance planning task changed across age groups. Second, we examined whether inhibitory control might be a key performance-limiting factor in the advance planning task by implementing two methodological alterations that have been demonstrated to improve children's performance in other tasks with inhibitory demands and analyzing their effect on performance measures.

### *Age-related trends in performance: Number of trials correct in each task*

There was no difference in overall performance (number of trials correct) based on gender in either of the tasks (Mann–Whitney $U$ test: sequential planning task, $U = 3361.0$, $N = 172$, $p = .29$; advance planning task, $U = 3134.5$, $N = 172$, $p = .08$), and so data for both genders were combined for all subsequent analyses. Within each age group, children succeeded in significantly more trials in the sequential planning task than in the advance planning task (Wilcoxon signed rank test: $Zs = -5.74$ to $-1.973$, $ps < .05$) (Fig. 3 and Table 1).

The order in which the two tasks were presented had a significant impact on performance in the sequential planning task for all age groups, with children getting significantly more trials correct when the sequential planning task was presented after the advance planning task (Mann–Whitney $U$ test: $Us = 548.0–738.5$, $ps < .05$). In the advance planning task, performance was better when this task was presented second for 6- and 7-year-olds ($U = 265.5$, $n = 60$, $p < .01$) and 9- and 10-year-olds ($U = 151.0$, $n = 52$, $p < .001$), but order of task presentation did not affect the performance of 4- and 5-year-olds in the advance planning task ($U = 340.0$, $n = 60$, $p = .09$).

### *Sequential planning task*

#### *Success in 1-, 2-, and 3-step trials*

The 9- and 10-year-olds were significantly more likely to succeed in at least one 1-step trial than both the 4- and 5-year-olds and the 6- and 7-year-olds (Fisher's exact test: $p < .05$ for both) (Table 1). The performance of 4- and 5-year-olds did not differ from that of 6- and 7-year-olds. The 9- and 10-year-olds were also more likely to succeed in 2-step trials than the 4- and 5-year olds (Fisher's exact test: $p < .05$) but not the 6- and 7-year-olds. The performance of 4- and 5-year-olds did not differ

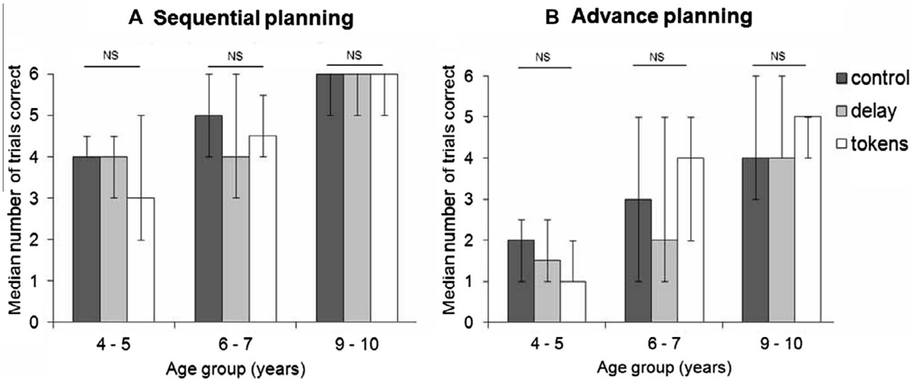

**Fig. 3.** Median number of trials correct for each age group as a function of experimental condition (control, delay, or tokens) in the sequential planning task (A) and the advance planning task (B). Error bars denote the inter-quartile range. NS indicates $p > .05$ in a Kruskal–Wallis test.

**Table 1**
Performance by age group for the two tasks.

| | 4–5 years | 6–7 years | 9–10 years |
|---|---|---|---|
| Overall trials correct (12) | 5.76 ± 0.29 | 7.57 ± 0.36 | 9.60 ± 0.24 |
| Sequential planning | | | |
| Trials correct (6) | 3.85 ± 0.17 | 4.38 ± 0.21 | 5.37 ± 0.12 |
| % ⩾1 × 1-step | 88.0 | 90.0 | 100.0 |
| % ⩾1 × 2-step | 85.0 | 93.3 | 100.0 |
| % ⩾1 × 3-step | 81.7 | 88.3 | 100.0 |
| Advance planning | | | |
| Trials correct (6) | 1.92 ± 0.21 | 3.18 ± 0.26 | 4.23 ± 0.22 |
| % ⩾1 × 1-step | 86.7 | 96.7 | 100.0 |
| % ⩾1 × 2-step | 26.7 | 50.0 | 76.9 |
| % ⩾1 × 3-step | 20.0 | 50.0 | 84.6 |
| % Start paddle only | 65.0 | 43.3 | 13.5 |

*Note.* "% ⩾1" variables represent the percentages of children succeeding in at least one of two 1-, 2-, and 3-step trials. Numbers in parentheses are the total numbers of trials presented. Values given are percentages or means ± 1 standard error.

from that of 6- and 7-year-olds. In 3-step trials, 9- and 10-year-olds were more successful than 4- and 5-year-olds (chi-square test: $\chi^2(1)$ values = 6.47–10.57, *ps* < .05). The performance of 4- and 5-year-olds did not differ from that of 6- and 7-year-olds. Within each age group, there were no significant differences in success rate between trials requiring a minimum of one, two, or three steps (*ps* > .05) (Table 1).

### *Advance planning task*

#### *Success in 1-, 2-, and 3-step trials*

The 4- and 5-year-olds were significantly outperformed by the 6- and 7-year-olds (chi-square test: $\chi^2(1) = 3.927$, $p < .05$) and the 9- and 10-year-olds (Fisher's exact test: $p < .01$) in 1-step trials (Table 1). The performance of 6- and 7-year-olds did not differ from that of 9- and 10-year-olds. There were significant differences between the performances of adjacent age groups in both 2- and 3-step trials, with 9- and 10-year-olds more likely to succeed than 6- and 7-year-olds (chi-square test: $\chi^2(1)$ values = 8.62–14.89, *ps* < .01) and with 6- and 7-year-olds being significantly more successful than 4- and 5-year-olds (chi-square test: $\chi^2(1)$ values = 11.08–11.87, *ps* < .01) (Table 1).

Within each age group, performance was significantly better in 1-step trials compared with 2-step trials (chi-square test: $\chi^2(1)$ values = 13.57–43.982, *ps* < .001) and in 1-step trials compared with 3-step trials ($\chi^2(1)$ values = 33.41–53.57, *ps* < .01) (Table 1). There was no significant difference in performance between the 2- and 3-step trials for any of the age groups (chi-square test: *ps* > .05).

The overall percentage of children only rotating the start paddle in all advance planning trials (i.e., never pre-positioning any paddles) differed significantly between age groups (Table 1), with 65.0% of 4- and 5-year-olds only ever turning the start paddle compared with 43.3% of 6- and 7-year-olds (chi-square test: $\chi^2(1) = 5.67$, $p < .05$). There was also a significant difference between the 6- and 7-year-olds (43.3%) and the 9- and 10-year-olds (13.5%) ($\chi^2(1) = 11.96$, $n = 112$, $p < .01$).

#### *Latency to first move*

Because of the experimenter-imposed 2-s delay in the delay condition, these trials were excluded from latency analyses and only data from control and tokens trials were included. Of this subset of trials, latency to first move was recorded for 656 of 688 trials in the sequential planning task and for 672 of 688 trials in the advance planning task. In the remaining trials it was not recorded accurately due to experimenter error (e.g., failure to press "start" or "stop" accurately on the stopwatch). These trials were excluded from the analysis. Log latency to first move for each trial was entered into an analysis of variance (ANOVA) with task (sequential or advance planning), condition (control or tokens), minimum number of steps (1, 2, or 3), and age group as between-trial factors as well as all possible interactions. There were significant main effects of all four between-trial factors (Table 2). Latency to first move was significantly longer in the advance planning task (5.008 ± 0.119 s) than in the sequential planning task (3.952 ± 0.083 s) (see Fig. 4) and was significantly longer in the control trials (4.704 ± 0.113) than in the tokens trials (4.263 ± 0.951 s). All pairwise differences among 1-, 2-, and 3-step trials were significant ($p < .001$ for all; see Fig. 4).

Following Bonferroni correction, there was no significant difference in latency to first move between 4- and 5-year-olds and 6- and 7-year-olds ($p = .09$), whereas all other pairwise comparisons revealed significant differences ($ps < .01$). There were two significant two-way interactions and no significant three-way interactions (Table 2). To interpret the interaction between condition and age group, ANOVAs were conducted to examine the effects of condition (control vs. tokens) on latency to first move within each age group. There was a significant main effect of condition on latency to first move for the 6- and 7-year-olds ($p < .01$), with shorter average latency in the tokens condition (4.022 ± 0.164 s) compared with the control condition (5.085 ± 0.221 s). Latency did not differ between conditions for 4- and 5-year-olds or 9- and 10-year-olds ($ps > .05$). To interpret the interaction between task and minimum number of steps, ANOVAs were conducted to examine the effects of task (sequential vs. advance planning) on latency to first move for each trial type (1-, 2-, and 3-step). Latency was significantly longer in the advance planning task than in the sequential planning task within each trial type ($ps < .05$); however, the difference was smaller between tasks in 3-step trials compared with 1- and 2-step trials.

**Table 2**
ANOVA of log latency to first move showing all main effects and all significant interactions.

| Source of variation | *df* | *F* | *p* |
|---|---|---|---|
| Main effects | | | |
| Task | 1 | 90.843 | <.001 |
| Condition | 1 | 36.071 | <.001 |
| Minimum number of steps | 2 | 97.241 | <.001 |
| Age group | 2 | 13.791 | <.001 |
| Significant two-way interactions | | | |
| Task * minimum number of steps | 2 | 3.767 | .023 |
| Condition * age group | 2 | 4.801 | .008 |
| Error | 1292 | | |
| Total | 1327 | | |

*Note.* Interactions not shown in the table were not significant ($p > .05$ for all).

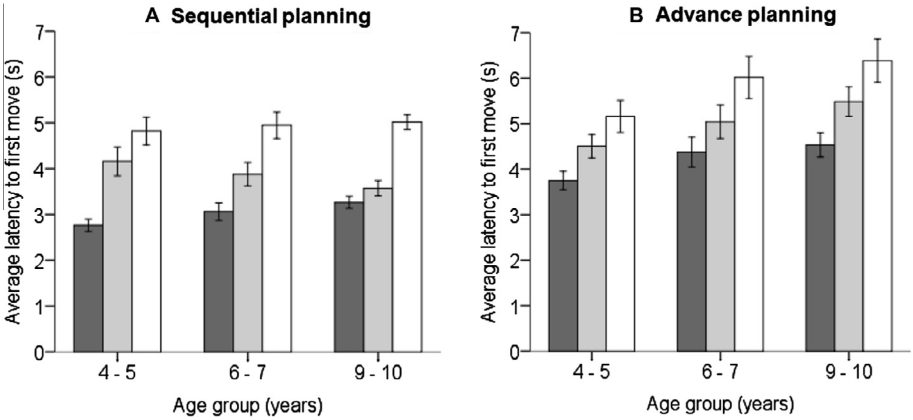

**Fig. 4.** Average (mean ± 1 standard error) latency to first move for each age group in 1-, 2-, and 3-step trials in the sequential planning task (A) and the advance planning task (B).

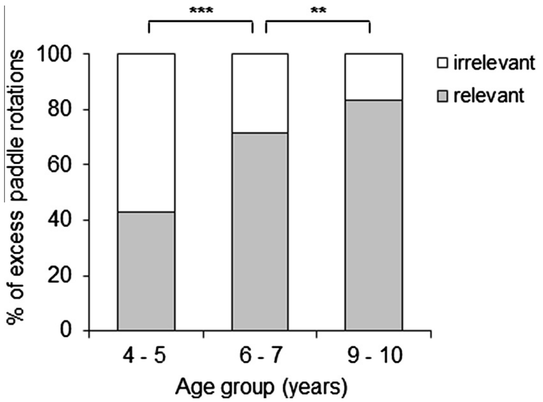

**Fig. 5.** Percentages of relevant and irrelevant excess actions in successful 2- and 3-step advance planning trials as a function of age group. Two asterisks (**) indicates $p < .01$, and three asterisks (***) indicates $p < .001$, in a chi-square test.

*Strategies for success in advance planning task*

Age group had a significant effect on how many excess actions (paddle rotations) were performed in successful 2- and 3-step advance planning trials (Kruskal–Wallis test: $\chi^2(2) = 15.09$, $n = 252$, $p < .01$). Post hoc pairwise comparisons revealed that 4- and 5-year-olds were significantly more likely to perform more excess actions in successful trials (4.24 ± 0.49) than 6- and 7-year-olds (2.14 ± 0.20) (Mann–Whitney *U* test: $U = 1025.0$, $n = 129$, $p < .01$) and 9- and 10-year-olds (2.20 ± 0.14) ($U = 1167.0$, $n = 152$, $p < .001$). The number of excess actions did not differ between 6- and 7-year-olds and 9- and 10-year-olds ($U = 4441.0$, $n = 219$, $p = .30$).

Of the excess actions performed by children in successful 2- and 3-step trials in the advance planning task, the percentages that were relevant to solving the task differed as a function of age group (Fig. 5).

Only 42.9% of 4- and 5-year-olds' excess paddle rotations were relevant to solving the trial compared with 71.8% for 6- and 7-year-olds (chi-square test: $\chi^2(1) = 37.77$, $n = 450$, $p < .001$) (Fig. 5). There was also a significant difference between the percentages of relevant paddle rotations made by 6- and 7-year-olds (71.8%) and 9- and 10-year-olds (83.3%) ($\chi^2(1) = 9.98$, $n = 524$, $p < .01$) (Fig. 5).

*Is inhibitory control a key performance-limiting factor? Impact of experimental conditions on performance*

Experimental condition (control, delay, or tokens) did not have a significant effect on the number of trials correct in any of the age groups in either the sequential planning task or the advance planning task (Kruskal–Wallis test: $\chi^2(2)$ values = 0.07–2.81, *ps* > .24) (Fig. 3). The percentages of children only rotating the start paddle did not differ significantly between conditions within any of the age groups (chi-square test: $\chi^2(2)$ values = 0.15–0.95, *ps* > .622).

## Discussion

Our primary aims were (a) to investigate the development of performance in the two paddle-box tasks across the three age groups tested and, hence, assess its suitability for a test of planning in children given that few appropriate tasks currently exist and (b) to better characterize the cognitive demands of the paradigm by establishing whether inhibitory control was a key factor limiting children's performance in the advance planning task.

*Development of planned problem solving: Age-related trends in paddle-box performance*

Children in all age groups performed significantly better in the sequential planning task than in the advance planning task. This was expected given the relative complexity of the two tasks. This finding also reflects the results of a previous study with apes (Tecwyn et al., 2013). The two tasks share some key demands, including the need to locate the item to be retrieved (start paddle), locate the open goal, and work out an appropriate sequence of paddle rotations to get the item from the start paddle to the goal. There are, however, some key differences. The sequential planning task is visuospatially simpler because all of the paddles are flat, which makes mentally visualizing the movement of the reward to the goal more straightforward. The item can also be controlled in a step-by-step manner. Therefore, it is possible to use a perceptually guided strategy (turn the paddle with the item on it toward the open goal) that incrementally moves the item toward the goal. This is true of 3-step ToL trials without any intermediate moves, which 4- and 5-year-olds are able to solve (Kaller et al., 2008). The advance planning task, on the other hand, is visuospatially more complex because it is not possible to follow a visual path from the reward to the goal given that relevant paddles need to be pre-positioned before the start paddle is rotated in order to create a valid path. This requires the anticipation of one's actions and planning in terms of searching ahead. This task can be likened to ToL trials with intermediate moves, which 4- and 5-year-olds struggle to solve (Kaller et al., 2008).

In both tasks, the average number of trials correct increased across age groups. In the sequential planning task, all age groups performed well in all trial types (1-, 2-, and 3-step). Even in the 3-step trials, 81.7% of 4- and 5-year-olds succeeded in at least one of two trials. Children within a given age group performed equally well across all trial types; that is, there was no evidence that they found 3-step trials to be more difficult than 1-step trials in the sequential planning task. This may be because it was possible to solve all of these trials using a perceptually guided strategy, which Kaller and colleagues (2008) suggested 4-year-olds were capable of using to solve ToL trials.

There was more variation in performance in the advance planning task across age groups, particularly in the 2- and 3-step trials. Only 20.0% of children in the youngest age group succeeded in at least one 3-step trial compared with 84.6% of 9- and 10-year-olds. In both 2- and 3-step trials, success increased across the age groups. Unlike in the sequential planning task, there was also variation between different trial types within each age group. Specifically, for all age groups, performance was significantly worse in 2- and 3-step trials compared with 1-step trials. This suggests that younger children found it difficult to look ahead and realize that other paddles needed to be rotated before turning the start paddle, as is thought to be the case for ToL trials with intermediate moves (Kaller et al., 2008). To gain a better understanding of why children were failing in the advance planning task, we examined the number of participants who only rotated the start paddle in every one of their trials (i.e., during the task they never rotated any paddles besides the start paddle). The percentage of children doing this decreased significantly with age, possibly due to the ongoing development of voluntary cognitive control (Asato et al., 2006). Furthermore, unlike older children, 4- and 5-year-olds performed badly in the advance planning task regardless of the order in which the two tasks were presented. It is possible that the youngest children had difficulty in "thinking outside the box" of the most obvious option (i.e., turning the paddle with the item on it) (Apperly & Carroll, 2009), whereas the older children were better able to overcome this, having had some experience (albeit limited) with the apparatus.

Average latency to first move was longer in the advance planning task than in the sequential planning task, which likely reflects the higher cognitive demands of the former (i.e., success in the sequential planning task required less preparation). Across the two tasks, latency to first move increased as the minimum number of steps required to retrieve the item increased. This is in keeping with findings from ToL studies (Albert & Steinberg, 2011; Luciana et al., 2009). The 9- and 10-year-olds had longer average latencies to first move than the two younger groups, which may be because older children are better at delaying immediate responding in favor of planning (Albert & Steinberg, 2011) due to the development of voluntary cognitive control of behavior that continues into adolescence (Asato et al., 2006). Increased preparation time with increasing age has also been reported in 4- and 5-step ToL trials (Albert & Steinberg, 2011; Asato et al., 2006; Luciana et al., 2009).

### *Development of planned problem solving: Strategies for success in advance planning task*

Because the advance planning paddle-box task can be solved in a variety of ways, it enables the examination of different strategies for success. The 4- and 5-year-olds performed significantly more excess paddle rotations (i.e., those that exceeded the minimum number of rotations in which the reward could be retrieved) when solving 2- and 3-step advance planning trials than children in the two older age groups. Asato and colleagues (2006) similarly found that excess moves decreased with increasing age in 4- and 5-step ToL trials. These findings reflect the fact that children in the youngest age group frequently solved 2- and 3-step advance planning trials by immediately setting up all of the paddles in a flat configuration and then retrieving the item in a step-by-step manner as in the sequential planning task (E. C. Tecwyn, personal observation, 2013). Older children, on the other hand, were more likely to pre-position paddles diagonally and, therefore, take fewer moves to solve a trial. These strategies clearly differ in the amount of planning required to achieve them but are, nevertheless, equally effective. Of the excess actions performed in successful trials, older children were more likely to perform excess actions that were relevant to solving the task as opposed to moving irrelevant paddles (i.e., paddles that could not influence the path between the item and the goal). This likely reflects older children's better understanding of the task, and greater efficiency in problem solving, possibly afforded by greater allocation of cognitive resources to planning or a switch to a strategy of planning ahead rather than one based on the use of perceptual cues (Kaller et al., 2008).

Currently, there is a lack of appropriate tasks for testing the planning skills of young children and infants (McCormack & Atance, 2011). The paddle-box provides a paradigm that may be more suitable than the currently available alternatives. Because the paddle-box was originally designed for use with apes, it has the benefit of not requiring participants to understand complex verbal instructions or adhere to specific rules, unlike in the ToL where participants need to understand, for example, that they are allowed to move only one disc at a time.

### *Is inhibitory control a key performance-limiting factor? Impact of experimental conditions on performance*

Neither of the two methodological alterations that were introduced to reduce the inhibitory demands of the advance planning task (imposing a short delay and replacing stickers with tokens) improved the performance of children in any of the age groups. Specifically, the experimental conditions did not reduce the likelihood of children performing the inappropriate response of always turning the start paddle immediately (causing the reward to become trapped), which we suggested was the prepotent response that needed to be inhibited to enable success in the advance planning task (Tecwyn et al., 2013). Although both of these measures have been demonstrated to reduce the likelihood of inappropriate prepotent responses in other tasks (Apperly & Carroll, 2009; Mitchell & Potson, 2001; Simpson et al., 2012), they are most effective in 3- and 4-year-olds, which is the age at which performance in standard inhibitory control tasks develops most, whereas by 5 years of age performance is at ceiling (e.g. Gerstadt et al., 1994). Although the noninhibitory demands of the paddle-box are different from those of the go/no-go task, windows task, and ToL, the inhibitory demands of all four tasks stem from the requirement for behavioral inhibition of a prepotent manual response as opposed to cognitive interference due to competing equivalent verbal responses, as is the case in the day–night task.

Therefore, it seems that inhibition of prepotent responding was not sufficient to enable success in the advance planning task. Albert and Steinberg (2011) found that prepotent response inhibition was not a predictor of ToL performance in 10- to 30-year-olds. The authors argued that as well as inhibiting immediate responding, participants also needed to sustain this delay in order to engage in effortful planning, and the development of this impulse control was reflected in the maturation of performance in complex ToL trials (Albert & Steinberg, 2011). Furthermore, Kaller and colleagues (2008) did not find any evidence that inhibitory control, measured using a version of the go/no-go task, explained the inability of 4-year-olds to solve ToL tasks requiring an intermediate move. However, whereas the ToL does not prime a strong prepotent response, the paddle-box does due to the salience of the desirable sticker. Therefore, the inhibitory demands of the advance planning task in this study are arguably higher than those of the ToL.

Other measures may have improved children's performance, including instructing them to plan their actions before starting and telling them to retrieve the item in the fewest possible moves (both of which improve ToL performance in adults) (Phillips, Wynn, McPherson, & Gilhooly, 2001; Unterrainer, Rahm, Leonhart, Ruff, & Halsband, 2003). In addition, asking children to indicate which paddles should be rotated rather than actually performing the paddle rotations may have enabled them to perform better. Hala and Russell (2001) found that 3-year-olds performed better on the windows task when they used a cardboard arrow to indicate which box their opponent should open compared with when they needed to point directly at the box with their finger. It is, however, questionable whether the youngest children in this study would have understood and followed such instructions for the paddle-box task, and we wanted to maintain the potential for comparability with nonhuman studies; therefore, we chose measures that could potentially be implemented in studies with other species. Delays have been implemented in nonhuman studies of problem solving, for example, by presenting the test apparatus behind a transparent Perspex barrier so that all of the components were visible but could not be manipulated and then removing the barrier after the delay period (e.g., Miyata, Gajdon, Huber, & Fujita, 2011). Tokens have also successfully been used in place of rewards in several studies with nonhuman primates (e.g., Evans, Beran, Paglieri, & Addessi, 2012).

### *Comparison with ape performance*

To our knowledge, there are very few direct comparisons of the planned problem-solving abilities of apes and children. Direct comparisons of the performance of different species using identical tasks are essential if we are to understand what distinguishes one species from another in terms of cognitive abilities (Dunbar, McAdam, & Connell, 2005). It is possible to draw some comparisons between the findings of the current study and Tecwyn and colleagues' (2013) findings with apes using the paddle-box, but caution is always required when making cross-species comparisons (Boesch, 2007). There are inevitably unavoidable differences in testing, including number of participants, number of trials, time frame of testing, testing environment, conspecific experimenter and verbal instructions for children, morphological differences, and reward type. However, designing tasks like the paddle-box that are operated via simple motor actions (turning paddles) and allow children and apes to interact with the apparatus in the same way (using their morphologically similar hands) maximizes the validity of comparisons. Hopper, Flynn, Wood, and Whiten (2010) further increased the validity of child–ape comparisons in their study by presenting children with the test apparatus inside a transparent box with access holes, thereby mimicking the barrier between apes and the apparatus that is formed by the mesh of their enclosure. The apes in Tecwyn and colleagues' (2013) study performed most comparably with the 4- and 5-year-olds in the current study. Both groups succeeded in the sequential planning task regardless of the minimum number of steps required to retrieve the item. Apes and 4- and 5-year-olds struggled in the advance planning task and in most trials only rotated the start paddle (i.e., failed to pre-position any paddles). In addition, unlike the two older child age groups, apes and 4- and 5-year-olds rotated a high proportion of irrelevant paddles. This is in keeping with findings from a recent study by Völter and Call (2014) that compared the performances of all four great ape species and 4- and 5-year-old children in a vertical maze task and found that young apes (<20 years) performed comparably to the children. Two younger apes even outperformed all of the 4-year-olds, demonstrating the ability to plan two steps ahead (one 5-year-old also succeeded in doing so), whereas the 4-year-olds were capable of planning only one step ahead (Völter and Call, 2014). Interestingly, the study revealed different limitations between the apes and the children; apes had difficulty in changing the direction in which they moved the reward on the different levels of the maze (motor control), whereas children struggled to shift their attention between subgoals of the task (Völter and Call, 2014).

## Conclusions

This study has demonstrated that the paddle-box is an appropriate tool for investigating the development of planning ability in children. The paradigm seems more suitable for conducting comparative investigations of planned problem-solving skills in children and apes than the currently available alternatives (e.g., ToL, route planning), particularly because sequential and advance planning can clearly be distinguished and task complexity can be systematically manipulated. There were clear developmental trends in performance, particularly in the advance planning task. Results from the sequential planning task suggest that, in keeping with other studies (e.g., Kaller et al., 2008; Miyata et al., 2009), the ability to plan a simple sequence of actions is present during early childhood, particularly where it is possible to succeed using a perceptually guided strategy. The capacity to plan a sequence with intermediate actions that requires looking ahead, as was required for success in the advance planning task, increases throughout childhood and is not mature by 10 years of age. Contrary to our predictions, inhibition of a prepotent response was not found to be a key performance-limiting factor for children in the advance planning task based on the methodological alterations that were implemented in this study. It is possible that it is not inhibition in terms of simply avoiding an inappropriate response that limits performance in the advance planning task but rather young children's (and apes') limitations in allocating sufficient cognitive resources to planning an appropriate response (Asato et al., 2006) or difficulty in "thinking outside the box" of the most obvious option.

## Acknowledgments

We thank the teachers and children at ARK Tindal Primary Academy, Clifton Primary School, and Stoneydelph Primary School for their help and participation. Thanks also go to Zoe Demery for assisting with data collection. We are grateful to Sarah Beck and three anonymous reviewers who commented on an earlier version of the manuscript. This study was funded by a Natural Environment Research Council studentship to E.C.T.

## References

Albert, D., & Steinberg, L. (2011). Age differences in strategic planning as indexed by the Tower of London. *Child Development, 82*, 1501–1517.

Anderson, V. (1998). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Neuropsychological Rehabilitation, 8*, 319–349.

Apperly, I. A., & Carroll, D. J. (2009). How do symbols affect 3- to 4-year-olds' executive function? Evidence from a reverse-contingency task. *Developmental Science, 12*, 1070–1082.

Asato, M. R., Sweeney, J. A., & Luna, B. (2006). Cognitive processes in the development of ToL performance. *Neuropsychologia, 44*, 2259–2269.

Baughman, F. D., & Cooper, R. P. (2007). Inhibition and young children's performance on the Tower of London task. *Cognitive Systems Research, 8*, 216–226.

Boesch, C. (2007). What makes us human (*Homo sapiens*)? The challenge of cognitive cross-species comparison. *Journal of Comparative Psychology, 121*, 227–240.

Diamond, A. (1990). Developmental time course in human infants and infant monkeys and the neural bases of inhibitory control in reaching. *Annals of the New York Academy of Sciences, 608*, 637–676.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology, 38*, 352–362.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). London: Oxford University Press.

Dunbar, R. I. M., McAdam, M. R., & Connell, S. O. (2005). Mental rehearsal in great apes (*Pan troglodytes* and *Pongo pygmaeus*) and children. *Behavioural Processes, 69*, 323–330.

Evans, T. A., Beran, M. J., Paglieri, F., & Addessi, E. (2012). Delaying gratification for food and tokens in capuchin monkeys (*Cebus apella*) and chimpanzees (*Pan troglodytes*): When quantity is salient, symbolic stimuli do not improve performance. *Animal Cognition, 15*, 539–548.

Gauvain, M., & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology, 25*, 139–151.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of 3½–7 year olds on a Stroop-like day–night test. *Cognition, 53*, 129–153.

Hala, S., & Russell, J. (2001). Executive control within strategic deception: A window on early cognitive development? *Journal of Experimental Child Psychology, 80*, 112–141.

Hopper, L. M., Flynn, E. G., Wood, L. A. N., & Whiten, A. (2010). Observational learning of tool use in children: Investigating cultural spread through diffusion chains and learning mechanisms through ghost displays. *Journal of Experimental Child Psychology, 106*, 86–97.

Hudson, J. A., & Fivush, R. (1991). Planning in the preschool years: The emergence of plans from general event knowledge. *Cognitive Development, 6*, 393–415.

Kaller, C., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural Brain Research, 216*, 1–8.

Kaller, C., Rahm, B., Spreer, J., Mader, I., & Unterrainer, J. M. (2008). Thinking around the corner: The development of planning abilities. *Brain and Cognition, 67*, 360–370.

Livesey, D. J., & Morgan, G. A. (1991). The development of response inhibition in 4- and 5-year-old children. *Australian Journal of Psychology, 43*, 133–137.

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in 4–8-year-old children. *Neuropsychologia, 36*, 273–293.

Luciana, M., Collins, P. F., Olson, E. A., & Schissel, A. M. (2009). Tower of London performance in healthy adolescents: The development of planning skills and associations with self-reported inattention and impulsivity. *Developmental Neuropsychology, 34*, 461–475.

McCormack, T., & Atance, C. M. (2011). Planning in young children: A review and synthesis. *Developmental Review, 31*, 1–31.

Mitchell, C. L., & Potson, C. S. L. (2001). Effects of inhibiting of response on Tower of London performance. *Current Psychology, 20*, 164–168.

Miyata, H., Gajdon, G. K., Huber, L., & Fujita, K. (2011). How do keas (*Nestor notabilis*) solve artificial-fruit problems with multiple locks? *Animal Cognition, 14*, 45–58.

Miyata, H., Itakura, S., & Fujita, K. (2009). Planning in human children (*Homo sapiens*) assessed by maze problems on the touch screen. *Journal of Comparative Psychology, 123*, 69–78.

Phillips, L. H., Wynn, V. E., McPherson, S., & Gilhooly, K. J. (2001). Mental planning and the Tower of London task. *Quarterly Journal of Experimental Psychology, 54*, 579–597.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The "windows task" as a measure of strategic deception in preschoolers and autistic participants. *British Journal of Developmental Psychology, 9*, 331–349.

Seed, A., Seddon, E., Greene, B., & Call, J. (2012). Chimpanzee "folk physics": Bringing failures into focus. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*, 2743–2752.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 298*, 199–209.

Simpson, A., & Riggs, K. J. (2007). Under what conditions do young children have difficulty inhibiting manual actions? *Developmental Psychology, 43*, 417–428.

Simpson, A., Riggs, K. J., Beck, S. R., Gorniak, S. L., Wu, Y., Abbott, D., et al (2012). Refining the understanding of inhibitory processes: How response prepotency is created and overcome. *Developmental Science, 15*, 62–73.

Tecwyn, E. C., Thorpe, S. K. S., & Chappell, J. (2013). A novel test of planning ability: Great apes can plan step-by-step, but not in advance of action. *Behavioural Processes, 100*, 174–184.

Unterrainer, J., Rahm, B., Leonhart, R., Ruff, C. C., & Halsband, U. (2003). The Tower of London: The impact of instructions, cueing, and learning on planning abilities. *Cognitive Brain Research, 17*, 675–683.

Völter, C., & Call, J. (2014). Younger apes and human children plan their moves in a maze task. *Cognition, 130*, 186–203.