



RESEARCH ARTICLE

Computational and behavioral markers of model-based decision making in childhood

Claire R. Smid¹ | Wouter Kool² | Tobias U. Hauser^{3,4} | Nikolaus Steinbeis¹

¹Department of Psychology and Language Sciences, University College London, London, the United Kingdom

²Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, Missouri, USA

³Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London, the United Kingdom

⁴Wellcome Centre for Human Neuroimaging, University College London, London, the United Kingdom

Correspondence

Claire R. Smid, Department of Psychology and Language Sciences, UCL, London, the United Kingdom.

Email: claire.smid.17@ucl.ac.uk

Funding information

European Research Council, Grant/Award Number: 715282; Sir Henry Dale Fellowship, Grant/Award Numbers: 211155/Z/18/Z, 211155/Z/18/B, 224051/Z/21; Jacobs Foundation, Grant/Award Number: 2017-1261-04; Medical Research Foundation, Grant/Award Number: 27023; Brain & Behavior Research Foundation; Philip Leverhulme Prize from the Leverhulme Trust, Grant/Award Number: PLP-2021-040; European Union's Horizon 2020 research and innovation program, Grant/Award Number: 946055; Max Planck Society

Abstract

Human decision-making is underpinned by distinct systems that differ in flexibility and associated cognitive cost. A widely accepted dichotomy distinguishes between a cheap but rigid model-free system and a flexible but costly model-based system. Typically, humans use a hybrid of both types of decision-making depending on environmental demands. However, children's use of a model-based system during decision-making has not yet been shown. While prior developmental work has identified simple building blocks of model-based reasoning in young children (1–4 years old), there has been little evidence of this complex cognitive system influencing behavior before adolescence. Here, by using a modified task to make engagement in cognitively costly strategies more rewarding, we show that children aged 5–11-years ($N = 85$), including the youngest children, displayed multiple indicators of model-based decision making, and that the degree of its use increased throughout childhood. Unlike adults ($N = 24$), however, children did not display adaptive arbitration between model-free and model-based decision-making. Our results demonstrate that throughout childhood, children can engage in highly sophisticated and costly decision-making strategies. However, the flexible arbitration between decision-making strategies might be a critically late-developing component in human development.

KEYWORDS

cost-benefit arbitration, decision making, metacontrol, model-based, model-free, reinforcement learning

1 | INTRODUCTION

To navigate our world successfully, we need to learn which of our actions lead to desirable outcomes. It is commonly theorized that reward-related learning in humans is guided by at least two decision-making systems that compete for control (Daw et al., 2005; Gläscher et al., 2010; Kahneman, 2003). One is a goal-directed and computa-

tionally costly model-based system, which can flexibly compare actions and their expected outcomes across contexts. The other is a habitual and computationally cheaper model-free system that ties rewards to specific cues and contexts, enabling the repetition of previously reinforced actions (Dickinson et al., 2002). The field of reinforcement learning provides a useful computational framework to dissociate contributions from these two systems to behavior (Daw et al., 2005; Dolan & Dayan, 2013; Gläscher et al., 2010). While model-based decision-making exploits the underlying hidden structure of an environment and

This work has not been published previously, is not currently under consideration for publication elsewhere, and the work has been reviewed and approved by all co-authors.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Developmental Science* published by John Wiley & Sons Ltd.



matches the rewards attained with the appropriate actions, model-free decision-making relies entirely on previously learned action-outcome contingencies. Although model-based decision-making can therefore be much more accurate, it comes at a cognitive cost. On the other hand, model-free decisions rely on previously learned action-reward outcomes and are therefore efficient, but cannot quickly adjust to changes in the environment. Optimally responding to different environmental demands, within the inherent processing limits of the human cognitive system, consequently requires dynamic arbitration between the costs and benefits of both decision-making systems (Lieder & Griffiths, 2020). For example, for everyday tasks, the efficiency of a model-free system might be preferred, while to be successful in novel or complex scenarios might require the more demanding but more accurate model-based system. Despite a wealth of studies showing that adults use both systems when making decisions, little is known about if and how these systems come to contribute to decision-making during human development.

From a young age onward, children are capable of making simple value-based decisions by learning which actions lead to positive, and which lead to negative outcomes. For example, even young infants have been shown to link actions and reward through gaze following (Ishikawa et al., 2020) and to learn the underlying hierarchical structure of a sequential decision-making task (Werchan & Amso, 2021). In addition, in a task where children were rewarded with cartoon video clips, preschoolers (3–4 years old) displayed action-outcome learning, by repeating actions that were rewarded in the past, and stopping certain actions when they no longer led to the same reward (Klossek et al., 2008, 2011). While these studies show that children can learn the relationship between their actions and subsequent reward, it is unclear whether children simply rely on model-free action-reward contingencies, or whether they can further employ this value-based learning to build an internalized model of the world, and use it to guide goal-directed behavior. Recent developmental studies using sequential decision-making tasks with 8–12-year-old children found no indication of contributions of a model-based system to choice before the age of 12 (Decker et al., 2016; Nussenbaum et al., 2020; Palminteri et al., 2016; Potter et al., 2017). Instead, the results from these studies suggest that the use of model-based decision-making strategies emerges in and increases through adolescence. These findings suggest that model-based decision-making might be a late-developing process, similar to other cognitive abilities such as fluid reasoning or inhibitory control (Otto et al., 2014; Potter et al., 2017).

Like many other studies investigating model-based decision-making in humans, these prior studies used a common sequential decision-making paradigm, often called the “two-step” task. Crucially, in the traditional version of the two-step task (Daw et al., 2011), using model-based decision making does not yield more reward than model-free decision making (Akam et al., 2015; Kool et al., 2016). In short, this is because the stochastic nature of the rewards and the transitions in the original two-step task make it difficult for a model-based system to effectively plan through the task structure (Kool et al., 2016). Indeed, recent variations of the traditional two-step task that simplified the transitional structure, which does allow a model-based system

Research Highlights

- Using both behavioral and computational markers, we find that children as young as five display model-based decision making, in contrast to previous developmental studies
- This means that in a reinforcement learning task, children can generalize information using an internalized model of the world
- However, children are not able to optimally arbitrate between decision-making strategies like adults, indicating that flexible control might be a late-developing skill
- This study sheds light on children's use of sophisticated decision-making strategies, proving that they can use similar constructs as adults

to outperform a model-free one, yielded a boost in model-based decision making in adults (Akam et al., 2015; Kool et al., 2016). Thus, the prior work reporting a lack of model-based decision making in 8–12-year-old children is unable to disentangle whether this reflected a general inability, or whether the stochastic task structure and lack of incentive stopped children from utilizing model-based decision making. Therefore, in the current work, we investigated whether children aged 5–11 years could engage in model-based decision-making when using a sequential decision-making task with a deterministic task structure that allowed for effective planning and greater incentives for using the model-based system.

In addition to a deterministic task structure, we used a further reward manipulation in the task to maximally incentivize the use of a model-based system. Previously, adults have been shown to increase their degree of model-based decision-making when greater rewards could be won (Bolenz et al., 2019; Kool et al., 2017; Patzelt et al., 2019). To date, it remains unclear whether or how children engage in effective and flexible metacontrol over distinct decision-making systems. Therefore, in addition to investigating whether children of this age range could engage in model-based decision making, we tested whether they arbitrated between model-free and model-based decision making in response to changes in the potential magnitude of reward. To this end, we used an environmental manipulation in the form of “high-stake” trials, where rewards were multiplied by a factor of five, and “low-stake” trials, where rewards were not multiplied. Optimal metacontrol on this task entails approximating the relative costs and benefits of using each system and increasing model-based decision making, which leads to higher rewards, for high-stake trials (Bolenz et al., 2019; Kool et al., 2017; Patzelt et al., 2019).

In sum, we address two questions; first, whether children aged 5–11 years can engage in model-based decision making using a novel sequential decision-making task; and second, whether children can demonstrate effective metacontrol over distinct decision-making systems. In contrast to previous findings, our results suggest that preadolescent children can engage in model-based decision-making, which



we demonstrate using both behavioral and computational methods. However, optimal metacontrol between goal-directed and habitual decision-making systems was not yet confidently expressed during childhood.

2 | MATERIALS AND METHODS

2.1 | Participants

Children were tested in pairs at a school in Greater London. Parental consent had been obtained prior to the study. Ethical approval for this study was obtained from UCL's Research ethics committee in compliance with UK national regulations. The present task was part of a larger battery of tests and was administered at the start of the battery. We used an a priori power analysis run in G*Power (Faul et al., 2007) to determine the sample size necessary to achieve similar power as in previous studies (Decker et al., 2016; Eppinger et al., 2013). Based on this, we determined that with a sample size of at least 60 children, we would achieve more than 90% power to detect a true age-related effect of comparable size (see [Supplementary Material](#) for the power analysis).

A total of 114 children were tested. Due to time constraints, some participants were not able to complete the entire task. We included children if they had (a) completed at least two thirds of the task, and (b) fewer than 30% missed trials. This led to the exclusion of 29 children (22 because of the task being cut short and seven because of missed trials). Missed trials were excluded from the analysis as participants did not receive rewards on these trials and therefore could not learn from them. On average, children missed 10% of the trials.

The final sample of children consisted of 85 participants (37 girls (44%), 48 boys). The mean age of children was 8.2 years ($SD = 1.6$), ranging from 5.0 to 11.4 years. Adult participants were tested at lab facilities at University College London. The adult sample consisted of 24 participants (11 females, (46%), 13 males), with a mean age of 25.2 years ($SD = 4.7$) ranging from 18.7 to 35.3 years. On average, adults missed 3% of the trials and none had to be excluded from the sample based on the two inclusion criteria described above. For further details on both samples, see the [Supplementary Material](#).

2.2 | Sequential decision-making task

2.2.1 | Task and narrative

We used a modified version of the novel task developed by Kool et al. (2017), which was designed to be more conducive to model-based decision making and to allow testing for the presence of metacontrol via low and high-stake manipulation that was more salient for children.

Participants were told that they were space explorers and that their mission was to collect as much treasure as possible from the two planets (red and purple) they could travel to. Each planet had one alien, which gave the participants treasure when they visited their planet. To be manageable for the younger children in our sample, our task consisted of 140 trials (compared to 201 trials in Kool et al., 2017). We conducted parameter recovery analyses of the current task with

100, 140, and 200 trials, to ensure that the model-based contribution (w) parameter had good recoverability for the trial numbers completed by participants in our sample. For these results, please see the [Supplementary Material](#).

Trials consisted of two stages. In the first stage, participants saw a pair of spaceships and had to choose one spaceship to travel to a planet. There were four spaceships in total and spaceships were always displayed in the same pairs, of which one spaceship always went to the red planet, and one spaceship always went to the purple planet, see [Figure 1a](#). In the second stage, participants had to collect treasure from the aliens on the planet. The amount of treasure that could be collected from each planet ranged between 0 and 9 treasure pieces and changed independently throughout the task following a Gaussian random walk with a standard deviation of 2, see [Figure 1b](#). Such drifting reward rates have been shown to promote learning and continuous monitoring of rewards won at each planet, in essence allowing a model-based system to capitalize on faster changes in rewards compared to the traditional two-step task (Kool et al., 2016; for full details on the task such as timings, see the [Supplementary Material](#)).

In this task, the difference between a model-based agent and a model-free agent is that a model-based agent can generalize between the spaceships that go to the same planet in each pair. For example, if the dark blue and the orange spaceship lead to the red planet, then a model-based agent should assign the same value to both spaceships. Thus, if a model-based agent chooses the orange spaceship, and receives a reward that is higher than expected on the red planet, the value of choosing both the dark blue and the orange spaceship increases, while for a purely model-free agent only the value of the orange spaceship increases. In short, the model-based agent generalizes reward experiences from one first-stage state (one pair of spaceships) to the other (other pair of spaceships) because they both lead to the same goal (the planet), whereas a model-free agent does not (Doll et al., 2015; Kool et al., 2016).

The current task was designed to encourage model-based decision-making by allowing a model-based agent to outperform the model-free agent in terms of reward gained throughout the tasks. This is accomplished due to the faster drifting reward rates, which a model-based agent can capitalize on by planning through an internal model of the task structure. This design leads to a positive correlation between the degree of model-based decision making and rewards earned, which was absent in previous versions of the task (see Kool et al., 2016 for a comprehensive overview).

2.2.2 | Stakes manipulation

To test whether our participants arbitrate between employing model-free and model-based systems depending on the rewards available, we employed low and high-stake trials. During the trials, participants received rewards in the form of pieces of blue space treasure. On a low-stake trial, the pieces of treasure won directly translated to the number of points won on that trial, for example, four pieces of blue treasure would have a value of four points, see [Figure 1c](#). In contrast, during a high-stake trial, rewards were multiplied by five; for example,

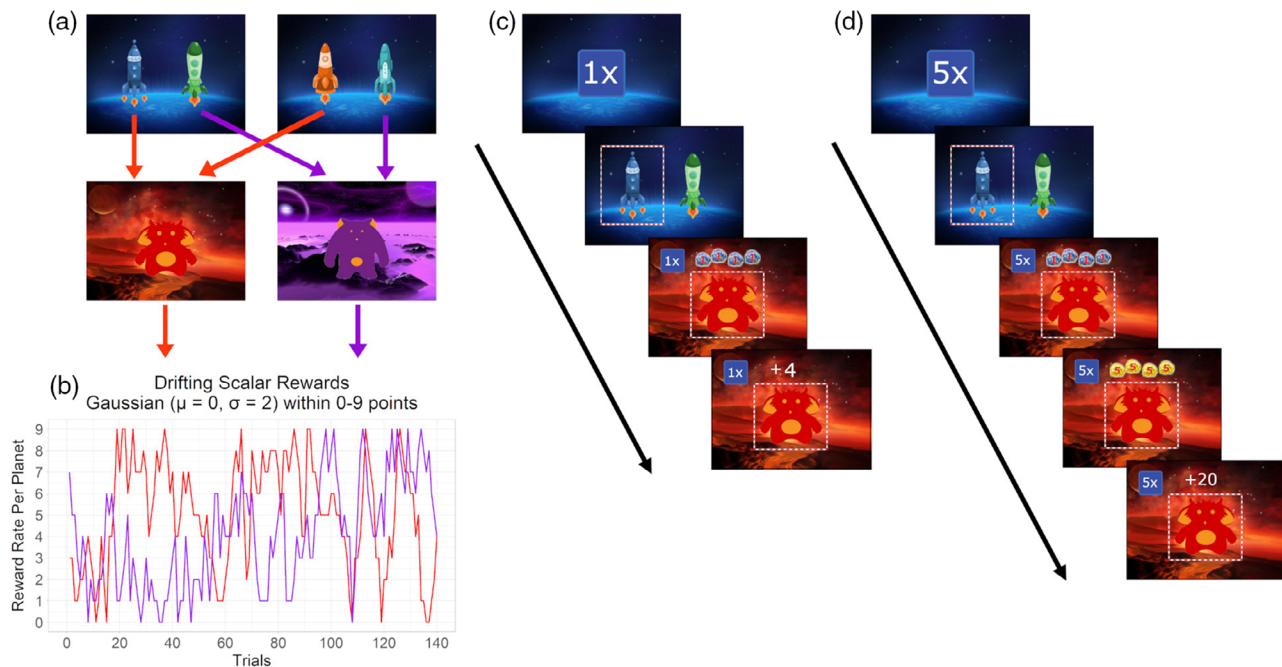


FIGURE 1 *Task Design.* (a) Schematic of the transition structure with arrows displaying deterministic transitions; if a participant chose the dark blue or the orange spaceship, they would always transition to the red planet. (b) At the planets, participants received rewards in the form of space treasure ranging between 0 and 9 pieces according to the drifting reward rate per planet. (c) At the start of the trial, participants saw the stake amplifier, which either showed “1x” for low-stake trials or “5x” for high-stake trials. Next, they saw a pair of spaceships and chose one after which they transitioned to either the red or the purple planet, where they had the opportunity to win pieces of treasure. During low-stake trials, pieces of treasure were displayed in blue with a red “1” on every piece, and participants received points equal to the number of treasure pieces shown. (d) During high-stake trials, the blue treasure was displayed first, and then, after a delay, turned into gold treasure with a red “5” on top of it, and the number of points received was multiplied by five

four pieces of treasure would have a value of 20 points. To make this difference between the stakes more salient for the children, on high-stake trials the treasure turned from blue to gold treasure after a short delay and displayed the number “5” in red on top of the gold treasure pieces, as opposed to “1” on the blue treasure for the low-stake trials, see Figure 1d. High- and low-stake trials were at an approximate 50/50 ratio and occurred randomly. For more details on the task and the stake condition, see our [Supplementary Material](#).

Metacontrol was calculated as a difference score in the degree of model-based decision-making expressed during the low- and high-stake trials. The degree of model-based decision making was measured via a weighting parameter, whereby a value closer to 1 indicated more model-based control, and a value closer to 0 as more model-free control. Using a model with two weighing parameters, one for each stake condition, we measured the difference in the values between the two parameters. A positive value indicated more model-based decision-making for high-stake trials and a negative value as more model-based decision-making for low-stake trials. A higher positive value reflects better metacontrol.

2.2.3 | Instruction phase

Before starting the main task, all participants completed an identical instruction phase, which took approximately 20 minutes. The main

task itself took approximately 25 minutes to complete. During the instruction phase participants learned (a) the deterministic transition structure (e.g., that one spaceship always went to the same planet; see Figure 1a), and participants were required to pass a criterion of four correct consecutive transitions to the red and purple planet respectively to continue the task; (b) that the amount of treasure changed over time (the drifting reward rates; see Figure 1b); (c) how to progress through a trial (e.g., first choose a spaceship, then collect treasure at a planet); and (d) the difference between high- and low-stake trials. This phase was identical for children and adults. No rewards were gained during the instruction phase and practice trials were not used for further analysis. For more details on the instruction phase, see the [Supplementary Material](#).

After the instruction phase, participants were told they would go on four missions to collect treasure during the main part of the experiment. Children were told that the more treasure they collected in the game, the bigger their present would be at the end of the study. Adults were told that for every 200 points, they would receive 50 cents (GBP).

We examined participants' understanding of the task by asking them to report the deterministic transition structure of the spaceships to the planets after the preparation phase. Due to missing data by tester omission, written responses from only 44 children were available. 80% of these children accurately reported the task structure. Of the 24 adults, 75% correctly reported where the spaceships went after practice.



There was no significant difference in the understanding of the task structure after the practice phase between children and adults, ($t(66) = 0.43$, $p = 0.670$, 95% CIs [-0.17, -0.26]), suggesting that the majority of the children learned the deterministic structure of the task.

2.3 | Statistical analysis and corrections

All statistical tests were conducted in R. For general effect sizes we report 95% confidence intervals and Cohen's d , and for regression results, we report the standard error of the mean (SEM). Cohen's d was acquired using the Effectsize package (Ben-Shachar et al., 2020). For t -tests, the default R Welch's t -tests were used, which do not assume equal variance across groups for an independent sample t -test, resulting in fractional degrees of freedom. When groups are compared for t -tests, the confidence interval reflects the 95% confidence of the mean difference between the groups. For correlations, the confidence interval reflects the 95% confidence range of values that contains the population correlation coefficient. For regression analyses, the package lme4 in R was used (Bates et al., 2015). When p -values are represented as "q," these "q-values" are multiple comparisons (FDR) corrected p -values using the default R STATS package. Dependent correlations were assessed using the COCOR package (Diedenhofen & Musch, 2015), and partial correlations were assessed using the PPCOR package (Kim, 2015).

We used an established dual-systems reinforcement learning model, which has been tested previously (e.g., Daw et al., 2011; Kool et al., 2016, 2017), to estimate the parameter solutions used to determine the degree of model-based decision making in the behavior of the participants. Model-fitting was conducted using the *mfit* package in Matlab (Gershman, 2016). In computational models, *priors* can be used which are values used to initialize the parameters of a model. If priors are kept "vague," they do not influence the parameter solution strongly, and only have a minimal effect on parameter solutions. Using priors helps with the accuracy of model-fitting, and we therefore used the same vague priors as used in a previous study investigating age effects in model-based decision making and metacontrol in aging adults (Bolenz et al., 2019; Gershman, 2016). We used Beta(2,2) priors for all model parameters bounded between 0 and 1 (learning rate (α), eligibility trace (λ), and the mixing weight(s) w), and a Gamma(3,0.2) prior for the inverse Softmax temperature (β), and for the two choice stickiness parameters (π and), ρ) we used Normal(0,1) priors (Bolenz et al., 2019). The model-fitting procedure we use to acquire our parameter solutions has the potential to introduce noise. To avoid this, we used model-free simulations to create a baseline to which we could compare the children (see Results). More details on the dual-systems reinforcement-learning model used for this study, the model comparisons, the model-fitting procedure, and the simulation procedure can be found in the *Supplementary Material*.

For the generalized linear mixed model, the package lme4 and the glmer command with family = binomial(link = "logit") were used

(Bates et al., 2015). The nested model selection was conducted using the AICcmodAvg package (Marc, 2020), and to visualize the plots, the ggeffects package was used (Lüdtke, 2018). For full details on the model comparison and approach, please see the *Supplementary Material*.

2.4 | Model-free simulation procedure

An important aim of this study was to investigate whether children in our sample showed influences of a model-based system in their behavior. However, since the model-based contribution parameter is bounded between 0 and 1, estimates of this parameter will always be larger (or equal) to zero. Meaning that noise in either the model-fitting procedure or in the behavioral performance of the participants can only push this parameter over the lower bound, and not under. We, therefore, created model-free simulations based on the estimated parameters solutions from the children (inverse temperature, learning rate, eligibility trace, and two choice stickiness parameters), but with the model-based contribution fixed to 0 to generate synthetic model-free behavior using the generative version of the dual-systems reinforcement learning model. Next, we used this synthetic model-free behavior to estimate a new model-based contribution parameter, which acted as our model-free baseline to compare the children to. For full details on the simulation procedure, please see the *Supplementary Material*.

All data, materials, and code for this paper are publicly available on Github: https://github.com/ClaireSmid/Model-based_Model-free_Developmental

3 | RESULTS

3.1 | Children perform above chance level and are not random

To assess whether children were sufficiently engaged with and capable of doing the task, we first compared their performance to chance level. Performance on the task was calculated as each individual's corrected reward rate, which reflected the average number of points a participant earned per trial, corrected for each participant's possible rewards based on the drifting reward rates (Figure 1b). This corrected reward rate tracks task performance against chance level (which was at 0). Scores lower than 0 indicate performance worse than chance, and scores higher than 0 indicate better than chance performance.

The mean corrected reward for children was significantly higher than chance ($t(84) = 3.20$, $d = 0.35$, $p = 0.002$, 95% CIs [0.003, 0.013]). Performance was also significantly correlated with age ($r = 0.32$, $p = 0.003$, 95% CIs [0.12, 0.50]). This suggests that the children were meaningfully performing the task, and that performance improved throughout childhood.



3.2 | Computational signatures of model-based decision making in children

The performance metric shows that children were generally able to perform the task. However, this above-chance level performance could arise from both successfully engaging a model-free or a model-based system. We thus investigated whether children specifically displayed model-based decision-making by fitting their behavior to an established dual-systems reinforcement-learning model (Daw et al., 2011; Gläscher et al., 2010). This model outputs several parameters that explain behavior (e.g., inverse temperature and a learning rate) and includes a weighting parameter that determines the relative contribution of each decision-making system to behavior, with weights close to 1 indicating a high degree of model-based decision making and weights close to 0 as mainly being model-free. As a higher value reflects a higher degree of model-based decision making, we will name this parameter “model-based contribution” throughout.

For both children and adults, we conducted a formal model comparison where we assessed four computational models, (1) a random model, (2) a simplified reinforcement learning model with three parameters (henceforth 3-parameter model), (3) a 6-parameter stake-agnostic dual-systems reinforcement learning model (henceforth 6-parameter model), (4) a 7-parameter metacontrol dual-systems reinforcement learning model with a model-based/model-free weighting parameter that was allowed to differ across stakes (henceforth 7-parameter model). We compared the models using k-fold cross-validation, Bayesian model selection, delta AICs, and parameter recoverability in two separate parameter recovery analyses, as well as a qualitative model assessment. From this comparison, the 6-parameter stake-agnostic dual-systems reinforcement learning model came out as the winning model overall. We fit the 6-parameter model to the data to assess model-based decision-making agnostic of stakes, and we use the 7-parameter model to assess metacontrol. For the full computational model, model comparisons, model-fitting details, and parameter recovery analyses, see the [Supplementary Material](#).

First, we investigated whether children displayed any model-based decision-making on the task over all trials combined. Children had an average model-based contribution of 0.52 (SD = 0.17), and given that this value is significantly larger than 0, ($t(84) = 27.40$, $d = 2.97$, $p < 0.001$ 95% CIs [0.48, 0.56]), it suggests that children used a model-based system during the task. However, because the model-based contribution parameter is bounded between 0 and 1, there is a possibility that noise (introduced during task performance or model fitting), could elevate the value of the model-based contribution to be greater than zero, even if the children only used model-free decision making.

To resolve this, we created model-free simulations based on the children's data. This resulted in a mean model-based contribution parameter of 0.28 (SD = 0.02) from these model-free simulations. Thus, a mixing weight value of 0.28 cannot be distinguished from pure model-free decision-making on the task and should be perceived as the baseline for testing the presence of model-based control. For full details on the simulation procedure, see the [Supplementary Material](#).

Critically, children's mean model-based contribution was in the 100th percentile of the model-free simulation's model-based contribution mean (100th model-free percentile: $w = 0.33$). This means that the mean of the children was larger than any mean value observed in the model-free simulations, indicating that children between 5 and 11 years of age show significant model-based decision making, ($t(84.22) = 12.47$, $d = 3.49$ $p < 0.001$, 95% CIs [0.20, 0.27]).

Additionally, we investigated whether the degree of model-based decision-making increased with age for the children. We found that there was a positive relationship between the degree of model-based decision-making and age ($r = 0.22$, $p = 0.042$), see Figure 2a.

Furthermore, we investigated whether the youngest children also showed significant model-based decision making. We conducted t -tests, separately for each year of age, correcting the p -values for false discovery rate. Every binned year of age showed a higher degree of model-based decision making than the model-free simulations, see Figure 2b (5-year-olds: $N = 7$, $t(6.00) = 4.28$, $q = 0.005$, $d = 10.36$, 6-year-olds: $N = 18$, $t(17.01) = 6.53$, $q < 0.001$, $d = 7.32$, 7-year-olds: $N = 15$, $t(14.00) = 5.21$, $q < 0.001$, $d = 7.11$, 8-year-olds: $N = 15$, $t(14.00) = 3.95$, $q = 0.002$, $d = 5.41$, 9-year-olds: $N = 17$, $t(16.00) = 4.47$, $q = 0.001$, $d = 5.62$, 10 ($N = 11$) and 11-year-olds ($N = 2$): $t(12.00) = 8.65$, $q < 0.001$, $d = 13.39$).

One of the main aspects of the current task design was that a higher degree of model-based decision-making leads to higher performance. To confirm this, we investigated the relationship between performance (the corrected reward rate) and the degree of model-based decision-making for the participants. Performance on the task was correlated to the degree of model-based decision making for the whole sample ($r = 0.51$, $p < 0.001$), showing that a higher degree of model-based decision making was significantly related to better performance on the task. This effect remained significant after controlling for age ($r = 0.37$, $p < 0.001$).

3.3 | Metacontrol of decision making for children and adults

In the current task, every trial is preceded by a “treasure amplifier” that indicates whether the current trial is a low or high-stake trial, see Figure 1c,d. During high-stake trials, any reward obtained on the trial is multiplied by five, while on low-stake trials, the reward is multiplied by 1 and therefore does not change in value. The current task entailed changes to a previously used task with adults (Kool et al., 2016, 2017) in the number of trials (140 as opposed to 201), the visualization of the stake condition, as well as a different testing environment (Amazon Mechanical Turk versus in-person testing). We therefore first tested whether we could replicate a stakes effect in an in-person adult sample. To investigate this, we fitted a reinforcement-learning model that included a model-based contribution parameter that differed for each stake condition to the adult data (Kool et al., 2017). There were thus two model-based contribution parameters, one for behavior during the low-stake trials and one for behavior during the high-stake trials. We

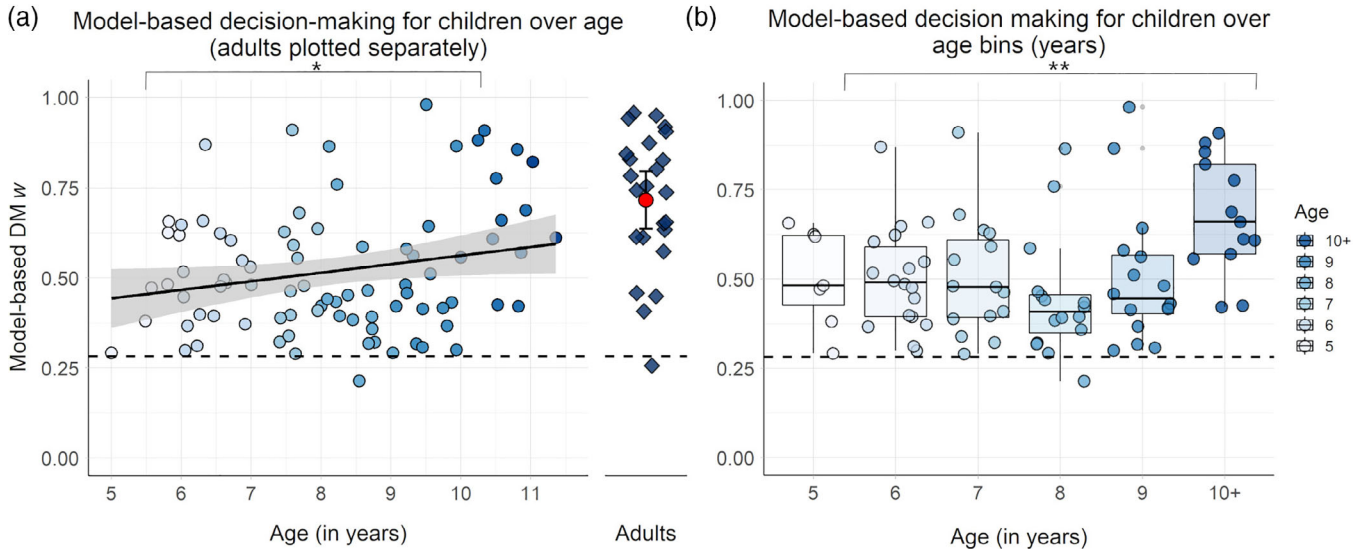


FIGURE 2 Model-based decision-making over age for children with the simulated model-free baseline. (a) The degree of model-based decision-making significantly increased with age for the children. The dashed line represents the grand mean of the model-free simulations, which acts as the simulated model-free baseline. The shaded area around the regression line represents the standard error of the mean. Adults are plotted separately. (b) Boxplots per rounded year of age for the children. As there were only two 11-year-olds, we combined these children with the 10-year-olds (10+). The dashed line represents the simulated model-free baseline. Asterisks indicate significance level, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. For Panel b, significance indicates the highest q -value of each binned year of age against the model-free simulations

conducted k -fold cross-validation to investigate whether both models could reliably predict choices made by the children and adults. Both models predicted behavior for children and adults significantly better than chance, but there was no significant difference in accuracy for either model (for details, see the [Supplementary Material](#)).

Adults showed a higher degree of model-based decision making during high-stake trials ($M = 0.71$, $SD = 0.19$), compared to low-stake trials ($M = 0.61$, $SD = 0.18$; $t(23) = 2.10$, $p = 0.047$, $d = 0.43$, 95% CIs [0.001 0.185]) see Figure 3a. This replicates previous findings of a stake effect on model-based decision making in adults (Bolenz et al., 2019; Kool et al., 2017; Patzelt et al., 2019).

Next, we assessed whether children's use of model-based decision-making was also affected by the rewards at stake. To investigate this, same as the adults, we fitted children's data to a reinforcement-learning model that included separate model-based contribution parameters for each stake condition (Kool et al., 2017).

Accordingly, we found no significant difference in model-based decision making between the low-stake ($M = 0.52$, $SD = 0.13$), and high-stake ($M = 0.52$, $SD = 0.13$) trials ($t(84) = -0.25$, $d = -0.03$, $p = 0.803$, 95% CIs [-0.03, 0.03]) for the children. This suggests that children did not show a stake effect like the adults, see Figure 3a.

When we compared children and adults directly, adults had higher model-based decision making than the children both during low-stake ($t(30.16) = -2.36$, $d = 0.65$, $p = 0.025$, 95% CIs [-0.18, -0.01]), and high-stake trials ($t(30.00) = -4.35$, $d = 1.21$, $p < 0.001$, 95% CIs [-0.27, -0.10]).

We next tested whether an effect of stakes on model-based decision-making might emerge with age for the children. Therefore, we

correlated the model-based contribution parameters for the low and the high-stake trials of the children separately with age and controlled the age-related slopes during high and low-stake trials for the correlation between the two contribution parameters. See Figure 3b for the age-related slopes over the two stakes. The difference between the slopes was not significant ($z = -0.50$, $p = 0.616$). We also plotted the group distributions and the differences in the individual participants' model-based decision making across the stakes, visualising the presence of a stakes effect for adults, and the lack of a stakes effect as a group for the children, see Figure 3c. Thus, a stakes effect was not apparent in the behavior of the children, suggesting that this ability may emerge later during development.

No other parameters (inverse temperature, learning rate, eligibility trace, or choice stickiness parameters) from the reinforcement-learning model were related to age for the children, see the [Supplementary Material](#).

3.4 | Behavioral signatures of model-based decision making for children and adults

To complement the computational modeling analyses, we used generalized linear mixed models to approximate a behavioral model-based decision-making measure, which was the probability of repeating a visit to a planet (stay probability) as a function of reward on the previous trial. We used the same regression method as in a previous version of the task (Kool et al., 2016). Using this method, the model-based component consists of a main effect of the previous reward on the probability

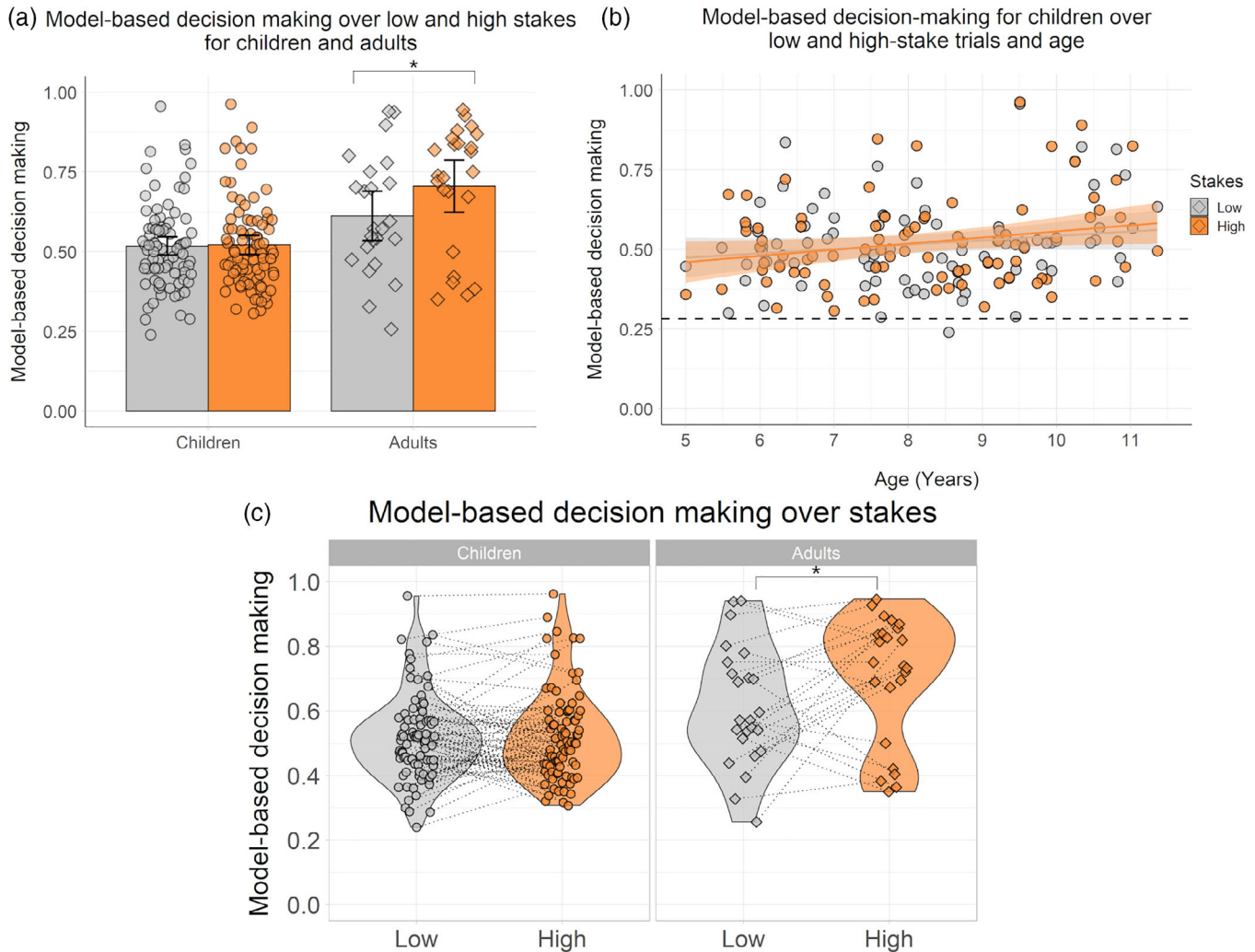


FIGURE 3 Model-based decision-making over stakes for adults and children. (a) Adults displayed a significantly higher degree of model-based decision-making for the high-stake trials, while children did not show a difference in the degree of model-based decision-making used over stakes. (b) this did not change over age for the children. The dashed line represents the model-free baseline. (c) connecting lines for participants' model-based decision-making across stakes plotted over the distributions for children and adults separately. Error bars depict 95% Confidence intervals, and shaded areas indicate SEM. Asterisks indicate significance level, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

of staying, whereas the reduced effect of previous reward when the starting state is different (compared to when it is the same) indicates a model-free component (Kool et al., 2016). Previous reward refers to the continuous points won by the participant on the previous trial and starting state similarity refers to whether the current starting state (the rocket pair) is the same as on the previous trial. The influence of previous reward on staying behavior approximates the transfer of experience from one starting state to the other, while the differential influence of previous reward on starting state similarity or difference can reflect a lack of transfer of experience between the starting states. Model-free and model-based systems should therefore generate different influences of starting state, as only the model-based system can effectively generalize over states, see Figure 4a.

First, we fitted an identical model to both children and adults that only looked at the influence of starting state similarity (whether participants saw the same spaceship pair as on the previous trial or the other

pair) and previous reward on stay behavior. For children, there was a main effect of previous reward on the probability to stay, indicating a model-based component ($\beta = 0.12$, $se = 0.02$, $z = 5.56$, $p < 0.001$). The interaction between previous reward and starting state similarity was not significant, showing that previous reward increased the probability to stay for both starting states similarly ($\beta = -0.003$, $se = 0.02$, $z = -0.14$, $p = 0.892$). In addition, there was a main effect of starting state ($\beta = 0.05$, $se = 0.02$, $z = 2.35$, $p = 0.02$). Thus, these results suggest that children could generalize successfully over starting states, and indicated a model-based component in their behavior, see Figure 4b.

For adults, there was also a main effect of reward on staying probability ($\beta = 1.09$, $se = 0.05$, $z = 22.81$, $p < 0.001$). There was no main effect of starting state ($\beta = 0.06$, $se = 0.05$, $z = 1.44$, $p = 0.149$), however, there was a small but significant interaction between starting state and previous reward ($\beta = 0.10$, $se = 0.05$, $z = 2.22$, $p = 0.026$), see Figure 4c. To be able to compare children and adults, we also

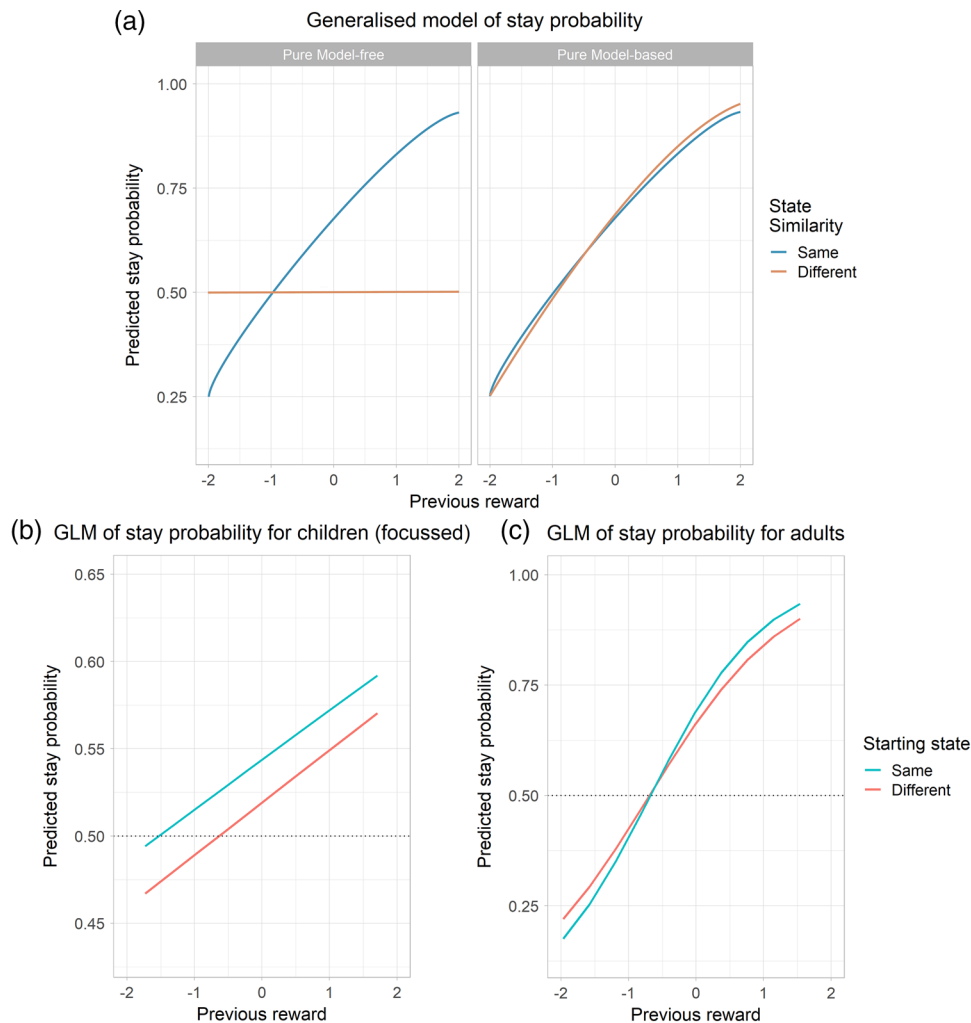


FIGURE 4 Model-free and model-based contributions to stay probability. Stay probability meant repeating a visit to the same planet (red or purple, see Figure 1a). (a) Examples of influences of pure model-free and model-based decision making on stay probability following previous reward. For a pure model-free system, stay probability only increases when the starting state (pair of spaceships) is the same. (b) Predicted results from a model investigating the influence of starting state. For children, across starting states, stay probability increased similarly with increasing previous reward, indicating a model-based effect. Note that the y-axis for children differs, as children generally showed a lower propensity to “stay.” (c) For adults, across the starting states the probability to stay also increased, indicating a model-based effect. The dotted lines for children and adults indicate the chance level of stay probability (50%). Continuous predictors in the models have been z-scored (e.g., Previous reward)

included groups in the models. the model-based predictor, previous reward, remains significant for the whole sample ($\beta = 0.12$, $se = 0.02$, $z = 5.55$, $p < 0.001$). We found that adults had a stronger effect of the model-based predictor on staying probability, indicated by an interaction between group and previous reward ($\beta = 0.98$, $se = 0.5$, $z = 18.67$, $p < 0.001$), as well as a higher probability to stay overall, based on a main effect of group ($\beta = 0.44$, $se = 0.10$, $z = 4.41$, $p < 0.001$). Adults also had a higher raw behavioral stay probability overall than the children, ($F(1,12631) = 120.9$, $p < 0.001$).

Thus, this suggests that adults also successfully generalize over starting states and that the effect of the model-based predictor was stronger for the adults than the children. The results from the regression models thus mirror the computational results. For further details on the regression models, see the [Supplementary Material](#).

3.5 | Best-fitting behavioral models for children and adults

Next, we conducted a nested model selection to find the best model to predict stay probability for both children and adults separately. In a previous logistic regression model, to more closely approximate the computational models, additional predictors were included alongside previous reward (the model-based component) and starting state similarity (same or different spaceship pairs). Namely, the difference in available rewards across the two planets on the previous trial (a proxy of reward history) and stake (high and low stakes), allows for investigating the influence of stake on choice behavior (Kool et al., 2016). For the current study, we also included age for the children. For both children and adults, we included a null model with only an intercept

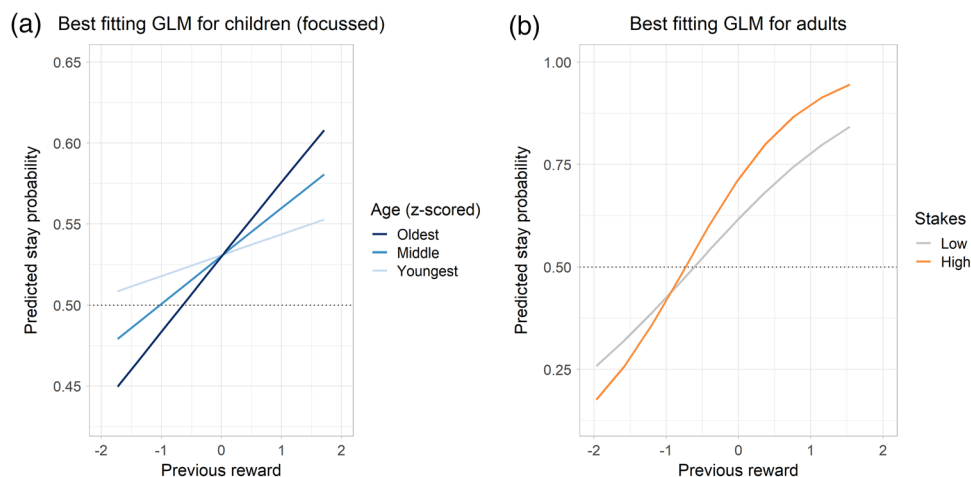


FIGURE 5 Best fitting generalized linear mixed models of stay probability for the children and adults. Stay probability meant repeating a visit to the same planet (red or purple, see Figure 1a). (a) Predicted results from the best-fitting model for children. Previous reward—the model-based component—was a significant predictor of stay probability, showing that children displayed model-based influences in the choice data. In addition, there was an interaction between previous reward and age (z-scored) showing that older children (Age z-scored = 1) showed a stronger increase in stay probability with reward than the younger children (Age z-scored = -1). Note that the y-axis for children differs, as children generally showed a lower propensity to “stay.” (b) For adults, previous reward was also a significant predictor, as well as stake. The interaction between previous reward and stake was also significant, showing that adults increased their stay probability during the high stakes for more reward. The dotted lines for children and adults indicate the chance level of stay probability (50%)

and no slope. For neither children nor adults was this null model the best fit.

For the children, the best-fitting model included previous reward (the model-based component) and age as fixed effects as well as their interaction (AIC weight (model probability) = 0.38; see [Supplementary Material](#)). Previous reward had a significant main effect on staying probability ($\beta = 0.12$, $se = 0.02$, $z = 5.60$, $p < 0.001$), while age was not a significant main effect ($\beta = -0.00$, $se = 0.04$, $z = -0.04$, $p = 0.967$), but the interaction between previous reward and age was significant ($\beta = 0.070$, $se = 0.02$, $z = 3.17$, $p = 0.002$), see Figure 5a. Thus, previous reward had a main effect on staying probability, indicating a significant model-based effect in the children’s choice behavior. The positive interaction with age shows that the influence of previous reward on staying probability increases with age.

For adults, the best-fitting model included previous reward, starting state and stake, as well as their interactions (AIC weight (model probability) = 0.83). There were significant fixed effects of previous reward (the model-based component) ($\beta = 1.14$, $se = 0.05$, $z = 22.78$, $p < 0.001$) and stake ($\beta = 0.22$, $se = 0.05$, $z = 4.88$, $p < 0.001$). Additionally, the interaction between previous points and stake was significant, indicating a stake effect ($\beta = 0.35$, $se = 0.05$, $z = 7.08$, $p < 0.001$), see Figure 5b. The interactions between previous points and state similarity was also significant ($\beta = 0.13$, $se = 0.05$, $z = 2.56$, $p = 0.010$), and the three-way interaction between previous points, starting state and stake ($\beta = 0.11$, $se = 0.05$, $z = 2.25$, $p = 0.025$), showed that there was a small effect for adults to be more likely to “stay” when the starting state was the same (same spaceship pair) during high stake trials.

Lastly, we tested whether using this approach we would also find that adults showed a higher degree of metacontrol than children.

We, therefore, fitted a model where we included group and stake as predictors, alongside the model-based (previous reward) and model-free (previous reward * starting state) predictors. The main effect of the model-based predictor remained significant, ($\beta = 0.12$, $se = 0.02$, $z = 5.54$, $p < 0.001$), and we saw that there was a significant three-way interaction between previous reward (the model-based indicator), stake and group ($\beta = 0.34$, $se = 0.05$, $z = 6.40$, $p < 0.001$), indicating that adults showed more model-based control during high stake trials. Thus, we see a stake effect repeated for the adults using the regression methods, and an absence of a stake effect for the children. This again mirrors the results from the computational models. For a full overview of the models and the results, see the [Supplementary Material](#).

4 | DISCUSSION

We investigated the development of model-based decision-making and how this is used adaptively across contexts in children aged 5–11 years. We report that when using a two-step task that encourages the use of computationally costly decision-making strategies, children aged 5–11 years demonstrated significant model-based decision making. This finding was supported by both computational and behavioral measures of model-based decision-making. Crucially, we found that even 5-year-old children showed robust model-based decision making, while the degree with which it was expressed increased further with age. However, whereas adults showed indicators of metacontrol by selectively increasing model-based decision-making for higher rewards, children did not. Combined, these findings demonstrate that children from as young as 5-years-old can engage in sophisticated decision-making



strategies on a sequential choice task, but that the optimal arbitration between strategies may be late-developing.

Our finding that children younger than 12-years-old display model-based decision making on a sequential decision-making task contrasts with prior studies reporting an absence of markers of model-based decision making before adolescence (Decker et al., 2016; Potter et al., 2017). These prior studies revealed a developmental increase in model-based decision making from childhood to adulthood, however, they also indicated that children as a group consistently showed signatures of model-free but not model-based decision making (Decker et al., 2016; Palminteri et al., 2016; Potter et al., 2017). In this study, using both computational and generalized linear models of choice behavior, the findings show that contributions of a model-based system to behavior are present before adolescence, and in children as young as 5-years-old. We attribute the discrepant findings between the current and prior work to task differences.

Compared to the original and commonly used two-step task (Daw et al., 2011), the present task encourages the use of model-based decision making by allowing a higher certainty in planning due to its deterministic transitions, and an increased rate of change in reward distributions (for an overview of all changes to incentivize model-based decision making, see Kool et al., 2016). The high complexity and uncertainty in the original two-step task, combined with the fact that more effortful model-based decision making did not lead to more rewards, may have hampered uncovering model-based decision making in children aged 8–12 years previously. Indeed, studies that employed an alternative two-step task with reduced transition complexity found increases in model-based decision-making in adults (Akam et al., 2015). It is not uncommon in developmental psychology that the removal of confounding variables and reduction of task complexity triggers competence shifts to younger ages (Scott & Baillargeon, 2017). Furthermore, our account is in line with previous findings of goal-directed behavior in infants and preschool-aged children in simple decision-making tasks (Klossek et al., 2008, 2011), showing that even very young children have the capacity to engage in sophisticated decision-making strategies when the task allows for this.

Contrarily, we found that, unlike adults, children did not prioritize model-based decision-making during high-stake compared to low-stake trials. Potentially, flexibly and swiftly arbitrating between decision-making strategies and anticipating which one is best suited to a certain situation might be the true late-developing skill (Nussenbaum & Hartley, 2019). For example, previous studies found that younger children are less aware of different environmental demands, and fail to respond to them proactively, for example by avoiding a more difficult condition (Chevalier, 2015; Niebaum et al., 2019). In addition, children, even up to late adolescence, might be less able to detect and assign values to relevant cues in the environment compared to adults, leading them to respond similarly to rewards of different magnitudes (Davidow et al., 2018; Insel et al., 2017). However, while the absence of metacontrol may reflect a genuine developmental effect in our sample, alternative interpretations are that children did not credit the high and low-stake conditions accurately enough or that the incentives used were not strong enough to uncover differences between the stakes

(Habicht et al., 2021; Veselic et al., 2021). Future work may wish to extend to using incentives that are even more salient to the present age group in order to establish whether metacontrol is genuinely absent in middle childhood. Another paper investigating the development of metacontrol in the form of prioritization of model-based decision making for high stakes over low stakes from adolescence to adulthood (ages 12–25) found that metacontrol continued to increase with age (Bolenz & Eppinger, 2021), but that in a sample between younger (ages 18–30) and older adults (ages 57–80), metacontrol declined for older adults (Bolenz et al., 2019). Thus, metacontrol might be particularly sensitive to developmental changes, peaking in early adulthood, and tapering off with advanced age. Exactly what drives this progression, for example, whether metacontrol is a unique stand-alone ability or whether it is reliant on executive functions or memory storage or manipulation, remains unclear.

While model-based decision-making was present throughout the age ranges in our sample, the display of model-based decision-making was still variable in this group and further increased with age. Individual differences in processes linked to model-based decision making, such as fluid reasoning, cognitive control, or working memory may well be able to account for an increase in the display of model-based decision making (Otto et al., 2013, 2014; Potter et al., 2017). Further research investigating such individual differences could shed light on the neurocognitive mechanisms underlying model-based decision-making in development. However, it remains important to consider the task context in which decision-making and cognitive control are studied (Plonsky & Erev, 2021), especially in developmental research.

When investigating the behavioral data, children showed a lower propensity overall to repeat a visit to the same planet, although the behavioral data indicated a higher probability to stay with higher previous reward, which indicates a model-based component in their behavior. The behavioral data lends itself to interpreting model-based decision making as it signals that starting state similarity did not lead to different behaviors of stay behavior similar to a pure model-free agent. Therefore, in their behavioral data, children also displayed that they generalized across starting states in the current task. However, our finding that children showed less overall likelihood to repeat a visit indicates one of the largest behavioral differences between children and adults. This might be due to children being less successful to exploit highly rewarding previous choices, or placing less importance on recent information, which is also reflected in their lower average values for inverse temperature and learning rate compared to adults (see the *Supplementary Material*). Thus, while children showed strong behavioral markers of model-based decision making in that their behavior did not differ across starting states, their behavior was different from adults, mainly due to being less likely to repeat visits to the same planet.

Additionally, we observed that children on average missed 10% of the trials, while adults missed 3%. While there were no differences in average reaction time between children and adults (suggesting the children were not at ceiling for responding), this could indicate that the 2-second response window for the first-stage state was fast for children of this age. Future studies might want to increase the

response window with the goal to limit timed-out trials for younger developmental samples.

Even though the current task is optimized to detect model-based decision-making compared to the Daw two-step task, it has less pronounced behavioral assessments of model-based decision-making. Future studies incorporating younger developmental samples may therefore also want to assess other two-step tasks that include a clear behavioral indicator of model-based control, for example, by using more conventional binary probabilistic rewards, and how this may change with age across childhood.

Lastly, while the dissociation between model-free and model-based decision making has been widely studied and supported (Bolenz & Eppinger, 2021; Bolenz et al., 2019; Doll et al., 2015; Gläscher et al., 2010; Kool et al., 2016; Kool et al., 2017; Otto et al., 2013, 2014; Patzelt et al., 2019), recent studies suggest that this dichotomy might be oversimplified, as well as potentially underestimating the ability of model-free control to approximate model-based control, for example via contextual learning or compound representations (Collins & Cockburn, 2020). Additionally, how distinct model-free and model-based prediction errors are in the brain remains under discussion, with some papers suggesting they might not be neurally distinct (Daw et al., 2011; Sanfey & Chang, 2008), and other studies reporting that distinct brain areas are involved for model-free and model-based prediction errors (Doll et al., 2015; Gläscher et al., 2010; Sambrook et al., 2018). Alternatively, new theories instead propose a more nuanced view of both reflexive habits and planning, combining them into a model that combines predictions about future events with flexibility following changes to rewards, dubbed successor representation (Momennejad et al., 2017). It seems likely that human decision-making is more complicated than a simple dichotomy of two opposing strategies that vie for control, and future models will likely become increasingly nuanced. However, in our current study, we believe that the dichotomy has aided us in understanding whether children aged 5–11 years old were able to apply an underlying transitional structure to their decisions and feel the current work is a valuable contribution to the field in including a wider range of developmental samples.

In summary, this study demonstrates the presence of sophisticated value-based decision-making strategies during childhood. We found that in a task where model-based decision making was tied to reward, and where the transitional structure was deterministic, children aged 5–11 years were able to engage in model-based decision making. The current study thus provides a crucial link between early goal-directed research on preschoolers and the computational modeling of model-based decision-making in adolescence. Interestingly, the ability to selectively amplify model-based decision-making during contexts with increased incentives was absent during childhood, indicating that metacognition, rather than model-based decision making, might be the cognitive process undergoing delayed development throughout childhood and adolescence. Future work spanning a range of paradigms, ages, and methodologies will be instrumental in charting the emergence and development of model-based control and its arbitration and link this to performance and competency-based developmental mechanisms.

ACKNOWLEDGMENTS

N.S. was supported by the European Research Council (European Research Council (ERC) grant agreement no. 715282, project DEVBRAINTRAIN), and a Jacobs Research Fellowship. T.U.H. is supported by a Sir Henry Dale Fellowship (211155/Z/18/Z; 211155/Z/18/B; 224051/Z/21) from Wellcome & Royal Society, a grant from the Jacobs Foundation (2017-1261-04), the Medical Research Foundation, a 2018 NARSAD Young Investigator grant (27023) from the Brain & Behavior Research Foundation, and a Philip Leverhulme Prize from the Leverhulme Trust (PLP-2021-040). He also has received funding from the ERC under the European Union's Horizon 2020 research and innovation program (grant agreement No 946055). The Max Planck UCL Centre is supported by the Max Planck Society. The authors thank Abigail Thompson, Keertana Ganesan, Sebastijan Veselic, Harriet Phillips, Somya Iqbal, and Ellina Guit for their help with data collection.

CONFLICT OF INTEREST

The authors whose names are listed on this paper certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

ETHICS STATEMENT

Ethical approval for this study was obtained from our university's research ethics committee in compliance with national regulations, project ID number: 12271/003.

AUTHOR CONTRIBUTIONS

Claire R. Smid conceived, designed, and performed the experiments and analyzed the data. Wouter Kool and Tobias U. Hauser helped analyze the data. Nikolaus Steinbeis conceived and designed the experiments. All authors wrote the manuscript and approved the final version before submission.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available on Github at https://github.com/ClaireSmid/Model-based_Model-free_Developmental

ORCID

Claire R. Smid  <https://orcid.org/0000-0002-0013-8814>

Wouter Kool  <https://orcid.org/0000-0003-3515-3982>

Tobias U. Hauser  <https://orcid.org/0000-0002-7997-8137>

Nikolaus Steinbeis  <https://orcid.org/0000-0001-8650-4725>

REFERENCES

Akam, T., Costa, R., & Dayan, P. (2015). Simple plans or sophisticated habits? State, transition and learning interactions in the two-step task. *PLoS*



- Computational Biology*, 11(12), 1–25. <https://doi.org/10.1371/journal.pcbi.1004648>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Ben-Shachar, M. S., Makowski, D., & Lüdtke, D. (2020). Compute and interpret indices of effect size. In *Cran* (p. <https://github.com/easystats/effectsize>). CRAN R Package. <https://github.com/easystats/effectsize>
- Bolenz, F., Kool, W., Reiter, A., & Eppinger, B. (2019). Metacontrol of decision-making strategies in human aging. *eLife*, 8, e49154. <https://doi.org/10.7554/eLife.49154>
- Bolenz, F., & Eppinger, B. (2021). Valence bias in metacontrol of decision making in adolescents and young adults. *Child Development*, 93(2), e103–e116. <https://doi.org/10.1111/cdev.13693>
- Chevalier, N. (2015). The development of executive function: Toward more optimal coordination of control with age. *Child Development Perspectives*, 9(4), 239–244. <https://doi.org/10.1111/cdep.12138>
- Collins, A. G., & Cockburn, J. (2020). Beyond dichotomies in reinforcement learning. *Nature Reviews Neuroscience*, 21(10), 576–586. <https://doi.org/10.1038/s41583-020-0355-6>
- Davidow, J. Y., Insel, C., & Somerville, L. H. (2018). Adolescent development of value-guided goal pursuit. *Trends in Cognitive Sciences*, 22(8), 725–736. <https://doi.org/10.1016/j.tics.2018.05.003>
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204–1215. <https://doi.org/10.1016/j.neuron.2011.02.027>
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–1711. <https://doi.org/10.1038/nn1560>
- Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From creatures of habit to goal-directed learners. *Psychological Science*, 27(6), 848–858. <https://doi.org/10.1177/0956797616639301>
- Dickinson, A., Wood, N., & Smith, J. W. (2002). Alcohol seeking by rats: Action or habit? Anthony. *The Quarterly Journal Of Experimental Psychology*, 55B(4), 331–348. <https://doi.org/10.1080/027249902440001>
- Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *Plos One*, 10(4), 1–12. <https://doi.org/10.1371/journal.pone.0121945>
- Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, 80(2), 312–325. <https://doi.org/10.1016/j.neuron.2013.09.007>
- Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D., & Daw, N. D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, 18(5), 767–772. <https://doi.org/10.1038/nn.3981>
- Eppinger, B., Walter, M., Heekeren, H. R., & Li, S. C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, 7(7 DEC), 1–14. <https://doi.org/10.3389/fnins.2013.00253>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology*, 71, 1–6. <https://doi.org/10.1016/J.JMP.2016.01.006>
- Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4), 585–595. <https://doi.org/10.1016/j.neuron.2010.04.016>
- Habicht, J., Bowler, A., Moses-Payne, M. E., & Hauser, T. U. (2021). Children are full of optimism, but those rose-tinted glasses are fading—Reduced learning from negative outcomes drives hyperoptimism in children. *Journal of Experimental Psychology: General*. Advance online publication, <https://doi.org/10.1037/xge0001138>
- Insel, C., Kastman, E. K., Glenn, C. R., & Somerville, L. H. (2017). Development of corticostriatal connectivity constrains goal-directed behavior during adolescence. *Nature Communications*, 8(1), 1–10. <https://doi.org/10.1038/s41467-017-01369-8>
- Ishikawa, M., Senju, A., & Itakura, S. (2020). Learning process of gaze following: Computational modeling based on reinforcement learning. *Frontiers in Psychology*, 11(March), 1–10. <https://doi.org/10.3389/fpsyg.2020.00213>
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, 93(5), 1449–1475. <https://doi.org/10.1257/000282803322655392>
- Kim, S. (2015). ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods*, 22(6), 665–674. <https://doi.org/10.5351/csam.2015.22.6.665>
- Klossek, U. M. H., Russell, J., & Dickinson, A. (2008). The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *Journal of Experimental Psychology: General*, 137(1), 39–51. <https://doi.org/10.1037/0096-3445.137.1.39>
- Klossek, U. M. H., Yu, S., & Dickinson, A. (2011). Choice and goal-directed behavior in preschool children. *Learning and Behavior*, 39(4), 350–357. <https://doi.org/10.3758/s13420-011-0030-x>
- Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When does model-based control pay off? *PLoS Computational Biology*, 12(8), 1–34. <https://doi.org/10.1371/journal.pcbi.1005090>
- Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological Science*, 28(9), 1321–1333. <https://doi.org/10.1177/0956797617708288>
- Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43, e1:1–60. <https://doi.org/10.1017/S0140525x1900061X>
- Lüdtke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, 3(26), 772. <https://doi.org/10.21105/joss.00772>
- Marc, J. M. (2020). AICcmoavg: Model selection and multimodel inference based on (Q)AIC(c) (R package version 2.3-1). <https://cran.r-project.org/package=AICcmoavg>
- Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, 1(9), 680–692. <https://doi.org/10.1038/s41562-017-0180-8>
- Niebaum, J. C., Chevalier, N., Guild, R. M., & Munakata, Y. (2019). Adaptive control and the avoidance of cognitive control demands across development. *Neuropsychologia*, 123(October 2017), 152–158. <https://doi.org/10.1016/j.neuropsychologia.2018.04.029>
- Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience*, 40, 100733. <https://doi.org/10.1016/j.dcn.2019.100733>
- Nussenbaum, K., Scheuplein, M., Phaneuf, C. V., Evans, M. D., & Hartley, C. A. (2020). Moving developmental research online: Comparing in-lab and web-based studies of model-based reinforcement learning. *Collabra: Psychology*, 6, 1–18. <https://doi.org/10.1525/collabra.17213>
- Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, 110(52), 20941–20946. <https://doi.org/10.1073/pnas.1312011110>
- Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2014). Cognitive control predicts use of model-based reinforcement learning. *Journal of Cognitive Neuroscience*, 27(2), 319–333. https://doi.org/10.1162/jocn_a_00709
- Palminteri, S., Kilford, E. J., Coricelli, G., & Blakemore, S. J. (2016). The computational development of reinforcement learning during



- adolescence. *PLoS Computational Biology*, 12(6), 1–25. <https://doi.org/10.1371/journal.pcbi.1004953>
- Patzelt, E. H., Kool, W., Millner, A. J., & Gershman, S. J. (2019). Incentives boost model-based control across a range of severity on several psychiatric constructs. *Biological Psychiatry*, 85(5), 425–433. <https://doi.org/10.1016/j.biopsych.2018.06.018>
- Plonsky, O., & Erev, I. (2021). To predict human choice, consider the context. *Trends in Cognitive Sciences*, 25(10), 819–820. <https://doi.org/10.1016/j.tics.2021.07.007>
- Potter, T. C. S., Bryce, N. v., & Hartley, C. A. (2017). Cognitive components underpinning the development of model-based learning. *Developmental Cognitive Neuroscience*, 25, 272–280. <https://doi.org/10.1016/j.dcn.2016.10.005>
- Sambrook, T. D., Hardwick, B., Wills, A. J., & Goslin, J. (2018). Model-free and model-based reward prediction errors in EEG. *Neuroimage*, 178, 162–171. <https://doi.org/10.1016/j.neuroimage.2018.05.023>
- Sanfey, A. G., & Chang, L. J. (2008). Multiple systems in decision making. *Annals of the New York Academy of Sciences*, 1128(1), 53–62. <https://doi.org/10.1196/annals.1399.007>
- Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, 21(4), 237–249. <https://doi.org/10.1016/j.tics.2017.01.012>
- Veselic, S., Smid, C. R., & Steinbeis, N. (2021). Developmental changes in reward processing are reward specific. *PsyArXiv*, 1–36. <https://doi.org/10.31234/osf.io/fzk9t>
- Werchan, D. M., & Amso, D. (2021). All contexts are not created equal: Social stimuli win the competition for organizing reinforcement learning in 9-month-old infants. *Developmental Science*, 24(5), e13088. <https://doi.org/10.1111/desc.13088>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Smid, C. R., Kool, W., Hauser, T. U., & Steinbeis, N. (2022). Computational and behavioral markers of model-based decision making in childhood. *Developmental Science*, e13295. <https://doi.org/10.1111/desc.13295>