Decisions, decisions, decisions: reinforcement learning, social, and patient decision-making in childhood

Claire Rosalie Smid

Supervisor: Prof. Nikolaus Steinbeis

Psychology and Language Sciences

University College London

Thesis submitted to UCL for the degree of Doctor of Philosophy in September 2022

I, Claire Rosalie Smid, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been properly cited and indicated in the thesis.

Signed:

Date:

Abstract

Human decision-making is the flexible way people respond to their environment, take actions, and plan toward long-term goals. It is commonly thought that humans rely on distinct decision-making systems, which are either more habitual and reflexive or deliberate and calculated. How we make decisions can provide insight into our social functioning, mental health and underlying psychopathology, and ability to consider the consequences of our actions. Notably, the ability to make appropriate, habitual or deliberate decisions depending on the context, here referred to as metacontrol, remains underexplored in developmental samples. This thesis aims to investigate the development of different decision-making mechanisms in middle childhood (ages 5-13) and to illuminate the potential neurocognitive mechanisms underlying value-based decision-making. Using a novel sequential decision-making task, the first experimental chapter presents robust markers of model-based decision-making in childhood (N = 85), which reflects the ability to plan through a sequential task structure, contrary to previous developmental studies. Using the same paradigm, in a new sample via both behavioral (N = 69) and MRI-based measures (N = 44), the second experimental chapter explores the neurocognitive mechanisms that may underlie model-based decision-making and its metacontrol in childhood and links individual differences in inhibition and cortical thickness to metacontrol. The third experimental chapter explores the potential plasticity of social and intertemporal decision-making in a longitudinal executive function training paradigm (N = 205) and initial relationships with executive functions. Finally, I critically discuss the results presented in this thesis and their implications and outline directions for future research in the neurocognitive underpinnings of decision-making during development.

Impact Statement

One's ability to make good decisions is essential in all facets of society. For example, understanding that one might need to forego small gains now in favor of security and larger gains in the future – for example, via the sacrifice of a portion of salary towards a pension contribution – allows individuals to plan for their future security and wellbeing. Throughout decision-making research, scientists have sought to uncover the facets that underlie human decision-making in the face of uncertainty and how these might be explained by normative, descriptive, and prescriptive theories. Crucially, through experimental research into human decision-making, researchers have come to understand that human decision-making is not purely rational or even consistent and susceptible to manipulations. An influential theory that sought to explain this perceived absence of rational decision-making discusses this from the perspective of bounded rationality. This concept states that human decision-making is rational, but it is constrained by the limits of the human mind, which potentially lie in individual differences in working memory capacity, attention, or cognitive control, abilities that are often captured under the umbrella term of executive functions.

Importantly, substantial changes in decision-making occur in childhood, where developmental studies observe that with age, we become more social, patient, and deliberate decision-makers. Coupled with significant changes in decision-making are changes in executive functions. From childhood to adolescence and adulthood, executive functioning generally improves, such that individuals become able to maintain and manipulate more information in working memory, have better impulse control over prepotent responses, and are better to switch between different tasks flexibly. Alongside this, critical structural changes occur in brain development, reflected in changes in cortical thickness. Thus, childhood is an ideal developmental period to investigate human decision-making and how this may be coupled to executive functioning improvements and brain anatomy changes. While the underlying neurocognitive mechanisms underlying reinforcement learning and social and intertemporal decision-making remain primarily unknown, inspecting these from a developmental lens provides us the opportunity to investigate the potential underlying relationships.

This thesis reports three key findings of decision-making in childhood. First, using a novel paradigm, I show that children are capable of using sophisticated goal-oriented decision-making strategies and can use the environment to their advantage, while previous studies suggested this skill only emerged in late adolescence. Second, I show that the ability to control when to flexibly exert effort in a reinforcement learning task for reward is linked to individual differences in cortical thickness and performance in inhibition tasks in childhood. Third, I investigate the plasticity of social and intertemporal decision-making in childhood via an executive functioning training paradigm and find that it did not lead to short-term or long-term training-related changes.

These findings make a significant contribution to the field of decision-making in development by carefully examining decision-making and linking it to individual differences in executive functions and brain anatomy.

Acknowledgments

Firstly, I would like to thank my supervisor Nikolaus Steinbeis for his guidance and for supporting me throughout my Ph.D. and the changes brought on by Covid-19. I am incredibly grateful to Niko for encouraging me to take new opportunities: to collaborate widely with other researchers, develop critical new skills and methods, and present my work whenever possible. I will be forever grateful for his guidance.

I would also like to thank the past and present members of the Developmental Change and Plasticity Lab for making my Ph.D. a great experience. I would especially like to thank Keertana Ganesan, Abigail Thompson, and Roser Cañigueral, who, throughout my Ph.D., were an incredible source of support through many in-person and virtual meetings. I also want to thank my fellow UCL students, who have helped me develop my passion for computational methods and together with whom I started the Machine Learning in Cognition at UCL (MLCog-UCL) academic society. It was inspiring to learn these new methods alongside them, and I am grateful for their enthusiasm and friendship; Christoffer Gahnstrom, Sarah Goodroe, and Nils Nyberg. I especially want to thank Sebastijan Veselic for being my main point of contact at UCL to run ideas by, check methods with, and a trusted co-author and friend. I am also incredibly grateful for the support I have received throughout our data collection from the Birkbeck-UCL Center for Neuroimaging (BUCNI) team and the many participants and their parents who took part in our studies.

In addition, I am grateful to my collaborators for the work reported in this thesis. Wouter Kool for his invaluable support in implementing the studies reported in Chapters 2 and 3, Tobias Hauser for his support and guidance and for allowing me to grow in my computational modeling skills, and Boris Bernhardt for his help with the structural magnetic resonance imaging analyses.

I would like to thank my partner Yihan for being with me for every step of this journey. I am forever grateful for his patience in listening to my ideas and long-winded conversations - and even when he didn’t want to - supplying me with helpful feedback on my clarity and logical arguments.

Finally, I would like to thank my parents for their endless and unconditional love and support and for providing me with the opportunity to pursue academic research. Their belief and confidence in my abilities have been priceless, and this thesis is dedicated to them.

Declaration for published works

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g., if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g., journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

<table>
<tr><td colspan="3">1. For a research manuscript that has already been published (if not yet published, please skip to section 2):</td></tr>
<tr><td>a) Where was the work published? (e.g., journal name)</td><td colspan="2">Developmental Science</td></tr>
<tr><td>b) Who published the work? (e.g., Elsevier/Oxford University Press):</td><td colspan="2">John Wiley & Sons Ltd.</td></tr>
<tr><td>c) When was the work published?</td><td colspan="2">22/05/2022</td></tr>
<tr><td>d) Was the work subject to academic peer review?</td><td colspan="2">Yes</td></tr>
<tr><td>e) Have you retained the copyright for the work?</td><td colspan="2">Yes</td></tr>
<tr><td colspan="3">[If no, please seek permission from the relevant publisher and check the box next to the below statement]:

☐ I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.</td></tr>
<tr><td colspan="3">2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3):</td></tr>
<tr><td>a) Has the manuscript been uploaded to a preprint server? (e.g., medRxiv):</td><td>Yes</td><td>If yes, which server? PsyArXiv</td></tr>
<tr><td>b) Where is the work intended to be published? (e.g., names of</td><td colspan="2">Click or tap here to enter text.</td></tr>
</table>

journals that you are planning to submit to)	
c) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.
d) Stage of publication	Please select.

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**
C.R.S. conceived, designed, and performed the experiments and analyzed the data. W. K. and T.U.H. helped analyze the data. N.S. conceived and designed the experiments. All authors wrote the manuscript and approved the final version before submission.
4. In which chapter(s) of your thesis can this material be found?
Chapter 2
5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g., if the paper was a single-author work):

Candidate:	Claire Rosalie Smid	**Date:**	26/08/2022
Supervisor/ Senior Author (where appropriate)**:**	Nikolaus Steinbeis	**Date:**	30/08/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

6. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
f) Where was the work published? (e.g., journal name)	Click or tap here to enter text.	
g) Who published the work? (e.g., Elsevier/Oxford University Press):	Click or tap here to enter text.	
h) When was the work published?	Click or tap to enter a date.	
i) Was the work subject to academic peer review?	Please select.	
j) Have you retained the copyright for the work?	Please select.	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
7. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
e) Has the manuscript been uploaded to a preprint server? (e.g., medRxiv):	No	**If yes, which server?** Click or tap here to enter text.
f) Where is the work intended to be published? (e.g., names of journals that you are planning to submit to)	Developmental Cognitive Neuroscience	
g) List the manuscript's authors in the intended authorship order:	Smid, C. R., Ganesan, K., Thompson, A., Cañigueral, R., Veselic, S., Royer, J., Kool, W., Hauser, T. U., Bernhardt, B. & Steinbeis, N.	
h) Stage of publication	Submitted	

8. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4**):**
C.R.S. conceived, designed, and performed the experiments and analyzed the data. K. G. and A. T. helped conceive the experiments and collect the data. J. R. and B. B. helped conceive the neuroimaging analyses, and W. K., T. U. H., the computational modeling. S. V. and R. C. helped revise and write the draft and analyses. N.S. conceived and designed the experiments. All authors read and approved the final version of the manuscript before submission.
9. In which chapter(s) of your thesis can this material be found?
Chapter 3
10. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g., if the paper was a single-author work):

Candidate:	Claire Rosalie Smid	**Date:**	30/08/2022
Supervisor/ Senior Author (where appropriate)**:**	Nikolaus Steinbeis	**Date:**	30/08/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

11. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**		
k) Where was the work published? (e.g., journal name)	Click or tap here to enter text.	
l) Who published the work? (e.g., Elsevier/Oxford University Press):	Click or tap here to enter text.	
m) When was the work published?	Click or tap to enter a date.	
n) Was the work subject to academic peer review?	Please select.	
o) Have you retained the copyright for the work?	Please select.	
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*		
12. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
i) Has the manuscript been uploaded to a preprint server? (e.g., medRxiv):	No	**If yes, which server?** Click or tap here to enter text.
j) Where is the work intended to be published? (e.g., names of journals that you are planning to submit to)	Nature Human Behavior, Nature Communications, Psychological Science	
k) List the manuscript's authors in the intended authorship order:	Smid, C. R., Ganesan, K., Thomspon, A., Canigueral, R. & Steinbeis, N.	
l) Stage of publication	Not yet submitted	
13. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)**:**		

C.R.S. conceived, designed, and performed the experiments and analyzed the data. K. G. and A. T. helped conceive the experiments and collect the data. K.G. conducted the confirmatory factor analysis, and R. C. helped write the manuscript. N.S. conceived and designed the experiments. All authors read and approved the final version of the manuscript before submission.			
14. In which chapter(s) of your thesis can this material be found?			
Chapter 4			
15. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g., if the paper was a single-author work):			
Candidate:	Claire Rosalie Smid	**Date:**	30/08/2022
Supervisor/ Senior Author (where appropriate)**:**	Nikolaus Steinbeis	**Date:**	30/08/2022

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

16. For a research manuscript that has already been published (if not yet published, please skip to section 2)**:**	
p) Where was the work published? (e.g., journal name)	Current Directions in Psychological Science
q) Who published the work? (e.g., Elsevier/Oxford University Press):	Sage Journals
r) When was the work published?	05/10/2020
s) Was the work subject to academic peer review?	Yes
t) Have you retained the copyright for the work?	Yes
[If no, please seek permission from the relevant publisher and check the box next to the below statement]: ☐ *I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.*	

17. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)**:**		
m) Has the manuscript been uploaded to a preprint server? (e.g., medRxiv):	Please select.	**If yes, which server?** Click or tap here to enter text.
n) Where is the work intended to be published? (e.g., names of journals that you are planning to submit to)	Click or tap here to enter text.	
o) List the manuscript's authors in the intended authorship order:	Click or tap here to enter text.	
p) Stage of publication	Please select.	
18. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4)**:**		

C. R. S, J. K. and N. S. all wrote the manuscript together.			
19. In which chapter(s) of your thesis can this material be found?			
Chapter 5 (General Discussion), section 5.9			
20. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g., if the paper was a single-author work):			
Candidate:	Claire Rosalie Smid	**Date:**	31/08/2022
Supervisor/ Senior Author (where appropriate)**:**	Nikolaus Steinbeis	**Date:**	30/08/2022

Table of Contents

Table of Figures

Table of Tables

Chapter 1. General introduction

1.1 A brief history of decision-making research: from rational to irrational creatures

In 2008, at the brink of the 2010 financial crisis, Wall Street guru Michael Burry realized that several subprime home loans in the US were in danger of defaulting. This information caused Burry to use over $1 billion of his investors' money to bet against the housing market, so-called "shorting", and to eventually make a fortune by taking full advantage of the impending economic collapse in America (Figure 1). A famous proverb in fiction and non-fiction applications is "Hindsight is 20/20", which means that it always seems easy to predict an event after it has occurred. Economists, researchers, and historians often claim to be able to predict the market. However, if the uncertainty of an event is immeasurable, it is nigh impossible to predict any event with certainty.

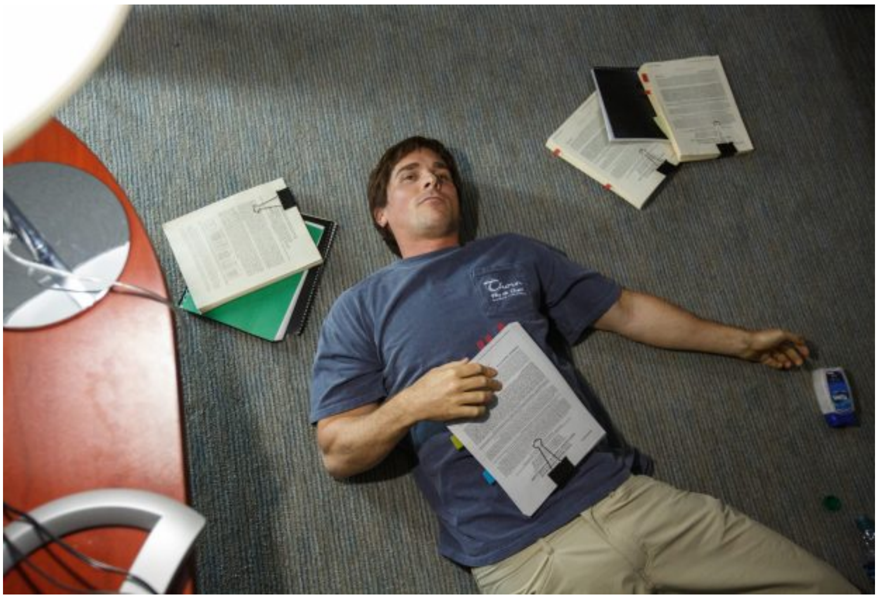


Figure 1. Christian Bale as Michael Burry in Paramount Pictures' *The Big Short*, 2015.

Despite this uncertainty regarding the future, people must make daily decisions that affect various domains of life. Some choices may have significant and long-term implications (e.g., deciding on a field of study, deciding to buy a house, deciding to have children, whom to marry). In contrast, other choices may be more short-term but still carry the potential for later implications (e.g., having unprotected sex, buying holiday insurance, or passing a vehicle on a two-way road). Therefore, understanding how we make decisions in uncertain contexts is crucial.

The emergence of decision-making as a formal field of study was intertwined with the birth of modern probability theory in the 17th century and has evolved significantly since then (Erez & Reyna, 2019). In the field of *judgment and decision-making*, choices listed above, whose outcomes are contingent upon whichever state of the world transpires, are classified as *decision-making under uncertainty* (Erez & Reyna, 2019). However, it should be noted that uncertainty about future events can sometimes be measurable and other times immeasurable. Thus, in some cases, we can calculate the available information into numbers that express the likelihood of observing certain events. For example, when rolling a die, we cannot say which number it will land on, but we can instead calculate the probability of obtaining the number six or an even number. However, the likelihood of uncertain events is incalculable in most real-life cases. For example, what is the probability of an economic crisis transpiring next year? Or what is our probability of being involved in a traffic accident in the next decade?

Decision-making research up until now has broadly fallen under three branches: *normative, descriptive, and prescriptive* analyses (Erez & Reyna, 2019). Traditional normative and descriptive decision-making models focused mainly on logic and cognition when modeling and predicting human choice behavior. Before the 1950s, normative theories led the way, where scientists believed that decision-making in the mind took place on a purely rational and mathematical basis, with the key assumption that the decision-maker was a logical, deliberative creature that obeyed basic rules of sound behavior (Elliott, 2019; Kacelnik, 1997). In short, normative theories draw from philosophical standpoints about how the ideal decision-maker *ought* to choose. This period gave rise to the Expected Value (EV) and the Expected Utility Theory (EUV), which introduced new and fundamental principles in probability theory and were pioneered by Pascal, Fermat, Cramer, and Bernoulli (Machina, 1987). EV captures the average across all possible outcomes weighted by their probabilities. For example, if we would roll a six-sided die six thousand times and average across all numbers, the EV would be close to 3.5 (Figure 2). The EV provided the ability to predict the average outcome of uncertain events accurately and was therefore seen as a guiding principle for rational-choice behavior. However, it soon became apparent that people do not consistently follow the EV principle in their decision-making. Thus, other factors must be at play in human decision-making.

$$E(X) = \mu = \sum x . P(x)$$

$$E[f(x)] = \sum_{D} f(x) \cdot p(x)$$

$$E(X) = \frac{(1+2+3+4+5+6)}{6} = \frac{21}{6} = 3.5$$

Figure 2. Expected Value of a six-sided die.

After *normative* theories of decision-making, which assume that the decision maker is a logical and deliberative creature, *descriptive* theories of decision-making became more pervasive. Descriptive analyses focus on how real, imperfect human beings make choices in practice, how they reason, and why they behave the way they do (Chandler, 2017; Kacelnik, 1997). In particular, it examines how human behavior differs from rational axioms derived from normative decision-making theories (Slavic et al., 1977). This field is, therefore, primarily based on empirical methods and statistical analysis conducted on choice behavior (Erez & Reyna, 2019). For example, it investigates how people's behavior and choices can be influenced and manipulated by introducing other factors in the decision-making context. Descriptive analyses are thus concerned with how real people *do* make choices.

Finally, the field of *prescriptive* decision-making can be seen as a mixture of both normative and descriptive analyses, with the primary goal to help people make better and more coherent choices (Slavic et al., 1977). Thus, prescriptive analyses offer decision-aiding tools in the form of rules and step-by-step guidelines to help people navigate their choices in a *normative* fashion, i.e., more rational and less biased by inconsistency, illogical decisions, or other biases (Erez & Reyna, 2019).

A pervasive theory for why humans did not adhere to rational, normative decision-making theories was introduced with the concept of “bounded rationality” (Simon, 1957). Herbert Simon was one of the first scientists to recognize that people’s rationality is limited and that the ideal decision-maker, as portrayed in normative models, thus could not replicate human decision-making. He argued that people utilize heuristics, which significantly simplify the decision-making process. Shortly before the conception of Simon’s theory, George Miller published his influential paper “The magical number seven, plus or minus two: Some limits on our capacity for processing information”, where he identified substantial limitations in human information processing (G. A. Miller, 1956). In short, he claimed that people could not exhaustively think about alternatives. Instead, decision-makers can remember and think about only a few chunks of information at a time (seven according to Miller, four according to some recent theorists (Erez & Reyna, 2019)), which limits or bounds their ability to make decisions.

Importantly, this links the ability to make rational decisions to other cognitive processes. This idea still holds today, and cognitive processes named *executive functions* (EFs) are thought to be critical to supporting flexible goal-directed behavior (Diamond, 2013). EFs capture cognitive abilities encompassing working memory, cognitive flexibility, and cognitive control in suppressing prepotent impulses. According to the concept of bounded rationality, rather than maximizing across all available alternatives, decision-makers aim at reaching a “good enough” criteria and then choosing the first option that reaches the threshold set by these criteria, so-called “satisficing”. Thus, people are limited but still rational and choose satisfactory but not necessarily optimal options.

As an example of this concept, imagine you are looking for a pub to have drinks in with four friends in London tomorrow. Clearly, the number of options to choose from is enormous, and the amount of relevant information to consider is exponentially higher: which location works for everyone? Do you want outdoor seating? How expansive is their drinks selection? Their food selection? Until what time do they serve food? Do they host pub quizzes? Many variables can be optimized (e.g., equidistant travel distance for all involved friends or multiple vegan food and drink options if a person in your group is vegan), and if you would attempt to pick the ultimate pub by solving this optimization problem, you might miss the proposed date and time for the drinks altogether. Thus, a common tactic used by people is to reduce the complexity of the problem, for example, focusing on familiarity and simply picking the pub you went to last week, which might cause some of your friends to travel for longer but allow you to come to a decision and propose a pub within an appropriate time limit.

In later works, deviations from rational decision-making (such as inconsistent choice behavior) were often attributed to humans’ bounded rationality (the brain's limited capacity to process information and perform complex calculations needed to maximize our well-being), as introduced by

Miller and Simon. The concept of bounded rationality, therefore, lays the groundwork for the link between decision-making and EFs.

1.2 A revolution in computation: formalized reinforcement learning models

Reinforcement learning is another branch of decision-related research formalized in the 1950s and has since experienced a dynamic history. Initially, two approaches to reinforcement learning existed independently; one started in animal psychology and trial-and-error learning. The other was rooted in engineering in the optimal control problem and how to solve this using value functions and dynamic programming. In the 1980s, both these approaches combined to constitute what we now know as modern reinforcement learning (Sutton & Barto, 2018).

Arguably the most crucial pioneer in trial-and-error reinforcement learning research in the early 1900s was Edward Thorndike, who coined the "Law of Effect". In his own words:

"Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction [to the animal] will, other things being equal, be more firmly connected with the situation so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort [to the animal] will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond." (Thorndike, 1911, p. 244)

Thorndike came to these conclusions via his experimental work, where he used puzzle boxes to study how animals learned. Thorndike placed a cat inside a puzzle box and then put a piece of meat outside the box. The boxes were enclosed but contained a small lever that, when pressed, allowed the cat to escape. He then observed the animal's approaches to escaping and obtaining the food, recording how long it took each cat to learn how to get out of the box. Eventually, the cats pressed the lever and opened the door so that they could eat the piece of meat. Even though the first interaction with the lever occurred by accident, the cats became likely to interact with it again in repeat experiments because they had received a reward immediately after performing the action. Thorndike observed that with each trial, the cats became faster at pressing the lever, opening the door, and obtaining the reward. Because pushing the lever had a favorable outcome, the cats were likelier to repeat the action (Thorndike, 1911).[1]

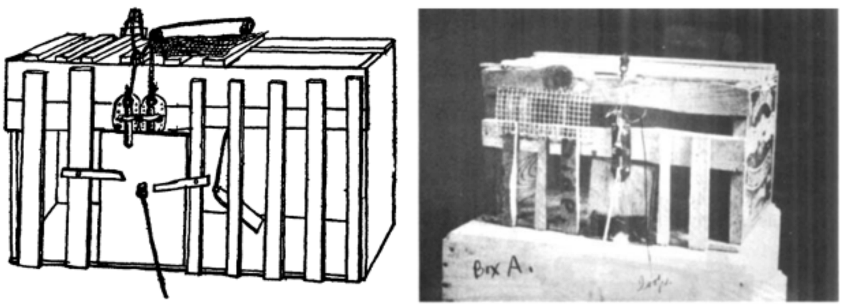


Figure 3. Thorndike's puzzle box.
Schematic (left) and photo (right) of Thorndike's puzzle box (Burnham, 1972).

This stipulation was coined the Law of Effect because it describes positive reinforcement on the tendency to select actions (Sutton & Barto, 2018).

[1] This research is similar to later work conducted by B. F. Skinner in operant conditioning.

This law describes two of the most important aspects of what is meant by trial-and-error learning. First, it is *selectional*, meaning that it involves trying alternatives, observing the outcomes, and selecting the most favorable option. Second, it is *associative*, meaning that the alternatives found by selection are associated with particular situations (Sutton & Barto, 2018). It is thus a simple way of combining search and memory; *search* by trying and selecting among many actions in each situation, and *memory* in the form of remembering what actions worked best and associating them with the situations in which they were the best option (Sutton & Barto, 2018). Importantly, this again links the field of decision-making research to EFs as supporting mechanisms to learn from previous rewards and punishments and to plan toward the subsequent desirable outcome.

In the other branch of reinforcement learning, scientists were concerned with the problem of "optimal control", which posits the problem of designing a controller to minimize a measure of a dynamic system's behavior over time (Sutton & Barto, 2018). It is concerned with finding the optimal path of all paths feasible for a system, for example, sending a rocket to the moon with minimal fuel consumption due to an optimized trajectory. One of the most influential approaches to this problem was developed in the mid-1950s by Richard Bellman and colleagues by applying the Hamilton-Jacobi equation from classical physics (R. Bellman, 1954; R. Bellman & Dreyfus, 1959; R. E. Bellman, 1957b). This approach used the concepts of a dynamical system's state and a value function or "optimal return function" to define a functional equation, which is now often referred to as the Bellman equation (Bellman, 1957). Dynamic programming became known as the class of methods that uses these equations to solve the optimal control problem.

The Bellman equation shows up everywhere in the field of reinforcement learning, being a central element of many reinforcement learning algorithms. In essence, the Bellman equation allows a value function to be decomposed in two parts, the immediate reward and the discounted future rewards. Thus, this equation allows a simplification of the computation of the value function, so that rather than summing over multiple time steps, the optimal solution of a complex problem can be found by breaking it down into simpler recursive subproblems and finding their optimal solutions (Figure 4). In simpler terms, the Bellman equation allows us to determine the value of the current state of the world and our action, and the next state of the world and action we might take.

Later stages of reinforcement learning research started merging the two branches of the trial-and-error learning approach rooted in animal psychology and engineering. This merging led to new computational models that assessed how an artificial agent could learn from the environment through trial-and-error and introduced new problems, such as the credit assignment problem (Sutton & Barto, 2018). This problem revolves around crediting success to the right actions when many different decisions were made that eventually led to success, which is essential to learning the right decisions to make to repeat that success (Minsky, 1961).

A basic reinforcement learning model revolves around an agent (an animal, human, or artificial agent) who takes certain actions in an environment and, depending on the state, receives an outcome, which could be a reward or punishment (Figure 4). The agent's goal in this problem is to find a sequence of actions that will provide them with a reward. Almost all reinforcement learning problems work by estimating value functions, which can be functions of a particular state of the world, or of a specific action while in a particular state (state-action pairs) (Sutton & Barto, 2018). A value function estimates how good it is for the agent to be in a given state (also written as V(s), where *V* represents the value and *s* the state) or how good a specific action is in a particular state (also written as Q(s, a) where *Q* represents the value, *s* the state and *a* the action). In short, the value of a particular state reflects the expected total reward that is obtainable while being in this state.

Besides the value function, another important function is that of state-action pairs, also called the action-value function or the Q-function (or just by Q). The Q-function defines the value of taking action *a* in state *s*, which can be written as Q(s, a) (or R(s, a)) (Sutton & Barto, 2018).

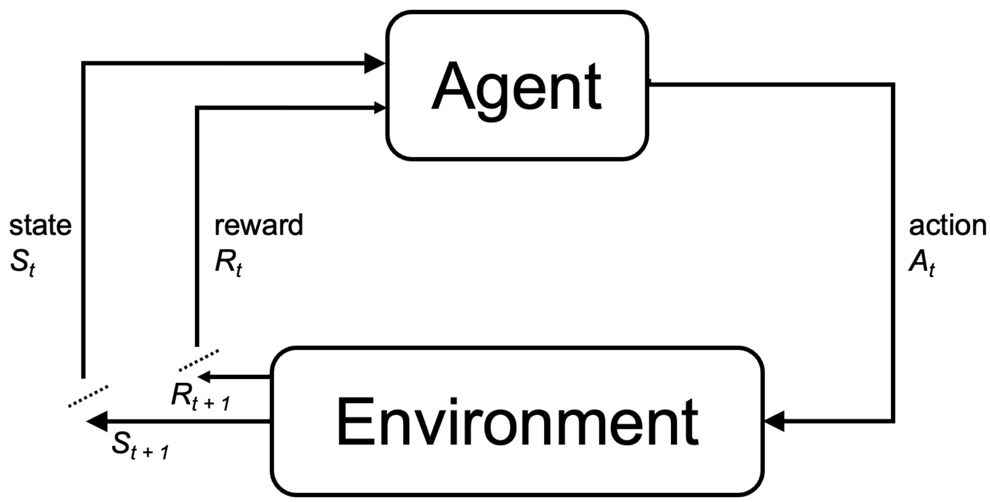


Figure 4. Reinforcement learning schematic.

To apply this to a real-world situation, let's use a contextual multi-armed bandit problem as an example[2]. Say we have a human agent, a boy named Timmy. Timmy is eight years old and visits his aunt Edna every month. Edna is generally happy, and when she is happy and Timmy shows her his grades, she sometimes gives Timmy extra pocket money. However, when she's not happy, she never gives Timmy extra pocket money. Over the years, Timmy learns that Edna is happy when it's sunny but becomes sad when it rains. Every time Timmy visits Edna, he has the choice to show her his grades or not. From his experience, Edna is happy when it's sunny, so the value of being at his aunt's when it's sunny has a higher value than being at his aunt's when it's raining. This reflects the value function, or the value Timmy has learned to associate with sunny weather. Now, Timmy has the option to show his grades or not. Thus, the Q-function for this situation can be written as +pocket money (sunny, show grades), which is positive because there is a probability he will receive a reward. In short, in this example, Timmy learns to take a certain action depending on the current state of the world that transpires, based on the previous reward he experienced.

[2] In a multi-armed bandit problem, the agent must maximize the sum of their rewards through a sequence of actions. It is so named after slot machines in a casino, where a gambler seeks to find the most rewarding slot machine to play on. In essence, the agent must learn what the optimal action is (e.g., in this example, whether to take the action (show the grades) or not (don't show the grades)).

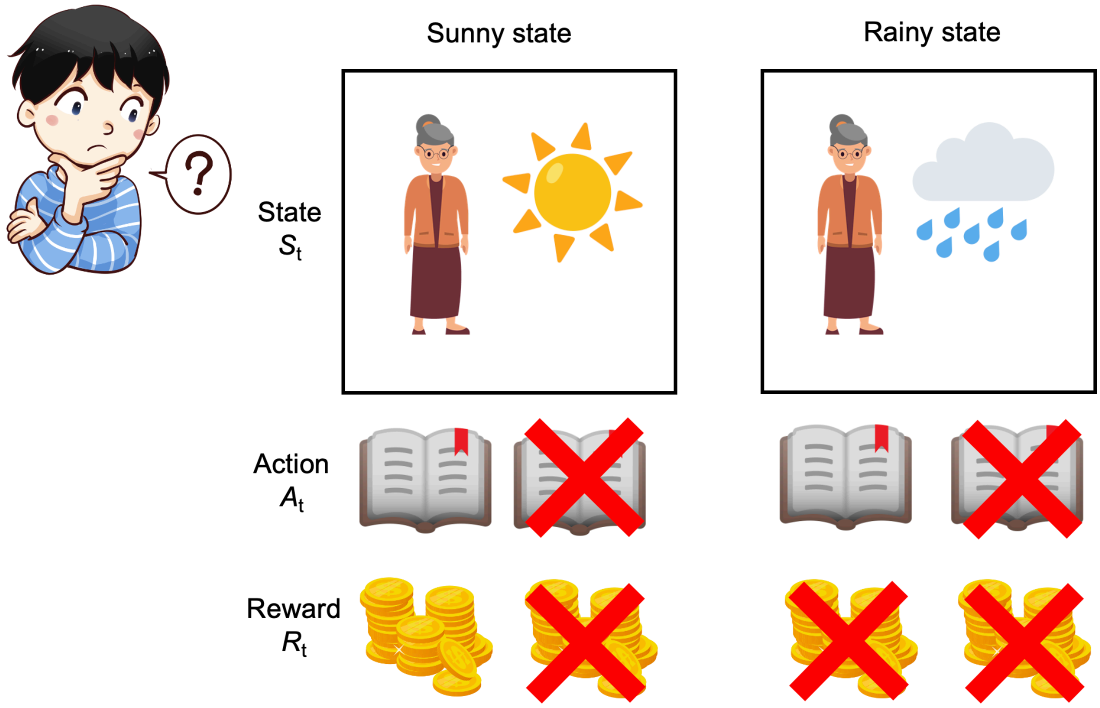


Figure 5. Contextual bandit example.
Depending on the context (weather: sunny, rainy), Timmy has to decide which action to take (show grades to aunt Edna, don't show grades to aunt Edna) to attain the reward (+pocket money, -pocket money).

Reinforcement learning models are pervasive because they allow researchers to capture and quantify specific components of (human) behavior. For example, in the above example, if we had logged all of Timmy's and Edna's interactions, the weather, and their outcomes, we could have fitted a model to capture how fast Timmy learned from his previous actions. In addition, we could have predicted which decision Timmy would be likely to make the next time he visited his aunt. In this thesis, I will use this approach to determine how children and adults rely on different strategies to learn from their past actions and make new decisions. The types of decision-making that are able to be dissociated with this approach will be discussed in the next section.

1.3 The theory of two minds: model-free and model-based

decision-making

The decision heuristics discussed in the first section of the introduction are colloquially referred to as habits. Habits are extremely important and effortless policies that can dictate our decision-making, which seemingly require little conscious thought and can be executed as if on autopilot (Dezfouli & Balleine, 2013; Dickinson, 1985; Dolan & Dayan, 2013). Let us consider the following example:

When commuting to work, you know which route to take without having to think about it consciously. You automatically walk to the tube station, habitually get off at the same stop, and walk to your workplace while your mind wanders. This is an effortless process. However, there is a tube strike today.

While your usual route to the tube was intuitive, you now find yourself spending some time planning alternative routes to work to select the quickest one. Are the buses still running? Which bus lines may get too crowded? Can you perhaps take a bike, or e-scooter, or walk there? How much will an Uber cost?

Our different responses to these processes demonstrate the distinctions between quick, intuitive, and habitual decisions and slower, goal-directed decisions, in short, a dualistic approach to decision-making (Daw et al. 2011). This dual-systems approach referred not only to decision-making but was also tied to the idea of cognitive abilities by psychologists Michael Posner and Charles Snyder in the 1970s in their book *Attention and Cognitive Control* (Posner & Snyder, 1975). They dissociated between cognitive processes which were either automatic or controlled, where automatic processes were characterized by four conditions:

1. They are elicited unintentionally
2. They only require a small number of cognitive resources
3. They cannot be stopped voluntarily, and
4. They happen unconsciously

Controlled processes were likewise characterized by four conditions:

1. They are elicited intentionally
2. They require a considerable amount of cognitive resources
3. They can be stopped voluntarily
4. They happen consciously

Similar to automatic processes, habits, and heuristics are both forms of model-free learning, where decisions are made based on experience and learned associations, rather than on a conscious understanding of the environment. Habits are actions performed automatically in response to specific stimuli, without conscious thought. Heuristics, on the other hand, are simple rules-of-thumb or mental shortcuts that allow for quick decisions in situations where a more deliberate, model-based approach would take too much time or mental effort. Model-based learning, in contrast, involves using a more explicit understanding of the environment and the relationships between actions and outcomes to make decisions. This approach is more flexible and adaptive than model-free learning, but also requires more computational resources and mental effort.

Importantly, the developmental aspect of the dualistic approach is extremely interesting. For example, do we start as quick, habitual individuals but progressively become able to “think slow”? Some current developmental studies seem to suggest that this is, in fact, the case and that we start as automatic, habitual decision-makers and only slowly develop to become controlled, goal-directed decision-makers (Davidow et al., 2018; Decker et al., 2016). This idea is supported by the fact that EFs that allow us to provide controlled responses, store and manipulate more information in our minds, and switch flexibly between tasks increase strongly throughout childhood (Buss & Spencer, 2014; Fiske & Holmboe, 2019; Ganesan & Steinbeis, 2022; Prencipe et al., 2011; Satterthwaite et al., 2013; Wiebe & Karbach, 2017). This, then again, leads us to the concept of bounded rationality. Is it the ongoing development of abilities that controls how many alternatives we can consider in our minds and how normative our decisions can become? Is childhood merely a period where we are unable to engage in what will become our peak rational decision-making once we have attained full brain maturation and executive control?

The theorized dichotomy between habitual and goal-directed forms of decision-making inspired a new blend of neuroscience-based reinforcement learning models. From a neuroscientific perspective, habits and goal-directed action systems appear to coexist in different corticostriatal circuits (Daw et al., 2011). While I mentioned in the previous section that these systems are thought to learn concurrently, they lead to different decisions as they link different rewards to different actions, as can be determined via q-learning reinforcement learning models (Balleine & O’Doherty, 2010; Dickinson, 1985; Kool et al., 2016; Lockwood et al., 2020). Computational approaches interpret these systems as two complementary mechanisms for reinforcement learning. The temporal-difference learning mechanism is associated with dopamine and reward prediction errors (RPE) and is model-free in the sense that it works by directly reinforcing successful actions rather than taking any underlying structure of the world into account (Daw et al., 2011). The goal-directed mechanism depends on a separate model-based reinforcement learning system, which works by using an internal model of the task to evaluate candidate actions (Beierholm et al., 2011; Daw et al., 2011; Doll et al., 2015; Kool et al., 2016). Thus, in theory, the choices recommended by model-free and model-based systems depend on their independent calculations. With good experimental design and accompanying models, the contributions of each system to decision-making are therefore possible to be dissociated.

If we revisit Timmy and his aunt Edna, Timmy has learned to associate the current weather with his aunt's happiness and, in turn, her happiness with his probability of receiving extra pocket money. In essence, Timmy has learned to apply a simple model of the world to predict the outcome of his actions. However, let's now consider that Timmy has a little brother, Howard. Howard also sometimes receives pocket money from his aunt Edna, but he has failed to make the connection between his aunt's happiness and the probability of pocket money. Instead, he only learned that by showing his aunt Edna his grades, he might receive pocket money. The next time Timmy and Howard visit Edna, it's raining, so Timmy decides to just enjoy some tea with his aunt and not to bother her with his grades. However, Howard confidently shows her his grades and solely receives a lukewarm verbal response in turn. Essentially, Howard has made a decision based on model-free decision-making (he only learned the action that may lead to reward), while Timmy has applied model-based decision-making (he used the current state of the world to plan his decision).

In 2011, Daw and his colleagues developed the two-step task, which is so named because each trial consists of two distinct steps (Figure 6). In this task, participants start in the first stage (pink rectangles) and pick one of the two options. If they pick the first option, 70% of the time, they will transition to the brown stage, where they will again need to pick between two options. For both brown options, they will receive a binary reward (0 or 1) according to a drifting probability bounded between 25% and 75% (Figure 6, bottom). However, 30% of the time, when selecting the first option in the pink stage, they will transition to the yellow stage instead. The yellow stage is otherwise identical to the brown stage in that they again need to pick between two options, where they will receive a binary reward according to the bounded drifting probability rate for reward for the two yellow stage options. However, as we can see from the drifting reward rates, sometimes one option is a lot better than the other options (in that the probability of receiving reward will be higher). Thus, participants need to continually assess how good the previous state and their action was.

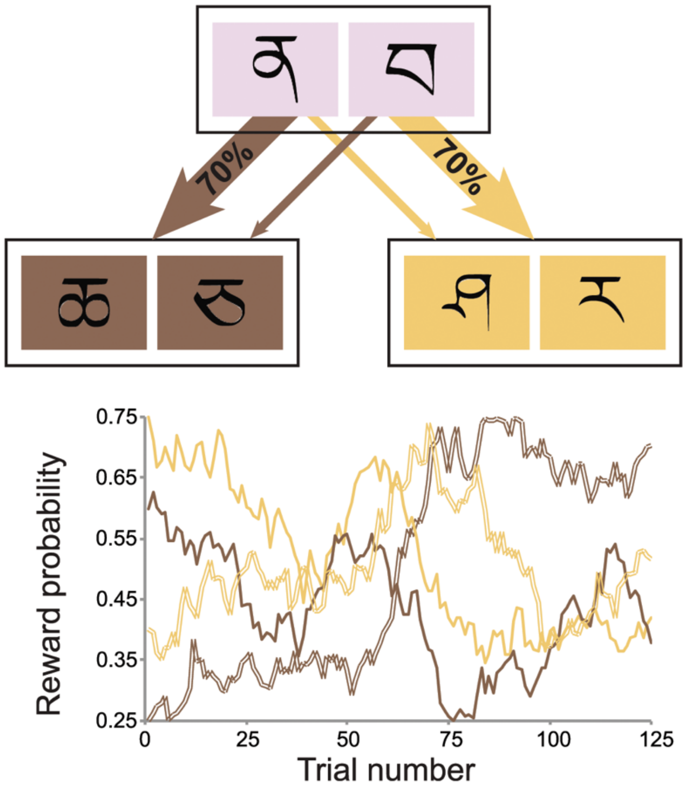


Figure 6. Two-step task structure and drifting reward probabilities. Recreated from Kool et al. 2016.

An attractive feature of this task is the stochastic transition structure, meaning that for the same action in the pink stage, 70% of the time (the thick brown and yellow arrows) will one action lead to the same second stage (common transition), but 30% of the time (the thin brown and yellow arrows) will it lead to the other (rare transition). This structure allows assessing how participants change their behavior following a rare transition and indicate whether they can effectively plan through the task transition structure. The four yellow and brown lines at the bottom of the figure represent the changes in reward probability (ranging between 25-75%) for the two brown and yellow second-state stages.

For example, a participant chooses the first pink option, expecting to transition to the second brown stage. However, they experience a rare transition and transition to the yellow second stage instead. Here they pick one of the two options and receive a reward. The difference between a model-free and model-based decision maker becomes apparent in their next decision. A model-based decision maker would be more likely to change their initial choice and instead pick the second pink option because it is more likely to transition to the yellow second stage, where they just received a reward (Figure 7b). However, a model-free decision maker would be more likely to choose the first pink option again because they do not use the internal structure of the task to plan their decisions (Figure 7a). Instead, they link their choice “pick the first pink option” to having received the reward, although it is transitioning to the yellow second stage that led them to the reward.

Therefore, the utility of the task structure in decisions reflects model-based decision-making, while ignorance of the task structure reflects model-free decision-making. The leads to the different systems calculating different values and probabilities for taking each action. To capture the extent to which participants used one system or the other, the reinforcement learning model used in Daw et al. 2011 incorporated a weighting parameter *w*, which when close to 1 reflects contributions from a pure model-based decision maker, and when close to 0 reflects contributions from a pure model-free decision maker. Empirical data from this task shows that people are hybrid decision makers and will display a mixture of the contributions of both these systems in their behavior (Figure 7).

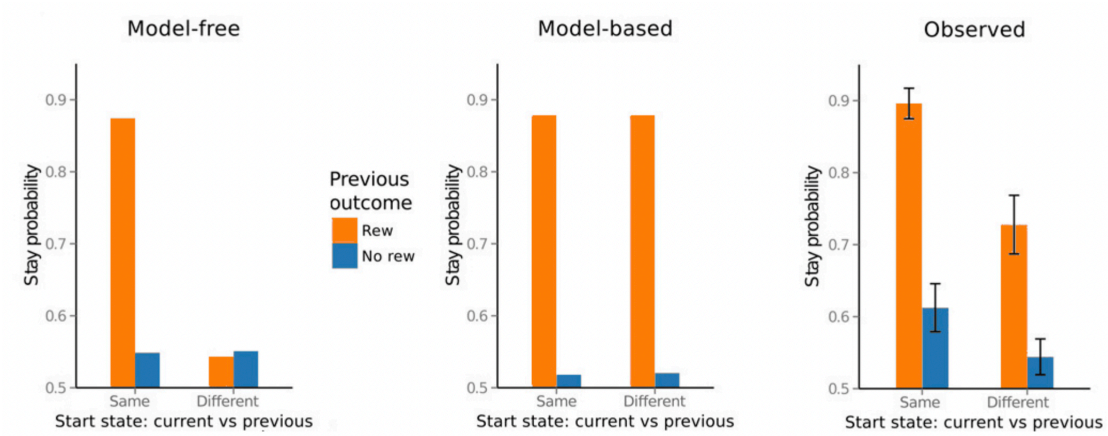


Figure 7. Probability of repeating the first-stage choice for three types of agents in a two-step task.

For model-free agents (a) the probability of repeating the previous choice is dependent only on whether a reward was obtained or not on the previous trial, not on the transition structure, or whether the previous transition was common or rare. Meanwhile, for model-based agents (b) this is reflected in an interaction between the previous transition type (common or rare) and previous reward, and the probability of transitioning to the same state (brown or yellow) where the reward was obtained. Behavioral performance on this task (c) reflects features of both model-based and model-free decision-making. Figure edited from Doll et al. 2015.

The idea of the model-based decision maker goes hand in hand with rationality and maturity. Since model-based decision-making was considered an advanced ability, it was considered to be an actual late-developing skill, one that only became available with finalized brain and behavioral maturation (Davidow et al., 2018; Decker et al., 2016; Nussenbaum & Hartley, 2019; Palminteri et al., 2016; Potter et al., 2017). In line with this theory, studies found that model-based decision-making seemed to be absent in children before 12 years of age and to become apparent and further increase during adolescence, peaking in young adulthood (Decker et al., 2016; Nussenbaum et al., 2020; Palminteri et al., 2016; Potter et al., 2017). However, as often found in developmental studies, this may not truly be the case (Scott & Baillargeon, 2017; Smid et al., 2022). In addition, the exact underpinnings of model-based decision-making, although theorized to be supported by EFs, remain unclear.

In the last few decades, empirical research has pointed out flaws in the dual-systems theories and added nuance to the strong dissociation proposed between them in earlier works (Dow, 1990; Momennejad et al., 2017). Current research suggests that while there may be a distinction between more intuitive and more deliberate forms of thinking and decision-making, they happen simultaneously, and all decisions are a mix of these systems (Daw et al., 2011; Feher & Hare, 2019; Gläscher et al., 2010; S. W. Lee et al., 2014; Sambrook et al., 2018). In addition, both systems can be influenced by biases and emotions and might not necessarily be rational (Stanovich & West, 2003). Even though this dichotomy between habitual, model-free, and goal-directed, model-based decision-making is now deemed much less divergent than previously thought, this proposed dichotomy fueled ground-breaking research into descriptive decision-making and which strategies people may rely on when making decisions. Adding to this, the field of reinforcement learning allowed new methods of quantifying the contributions of different systems to decision-making: formal models of habitual and goal-directed decision-making would create different reinforcing values for certain actions under different circumstances and thus, allow the contribution of either system to decision-making to be dissociated (Daw et al., 2011; Doll et al., 2015; Kool et al., 2016).

Prior work on the neural underpinnings of model-free and model-based decision-making has sought to uncover distinct signatures of associated prediction errors. Some studies suggest distinct regions for model-based decision prediction errors, such as the posterior parietal cortex (O'Doherty et al., 2015), the dorsomedial prefrontal cortex (PFC) (Doll et al., 2015), and the (dorso) lateral prefrontal cortex (DLPFC) in particular (Beierholm et al., 2011; Cremer et al., 2021; Doll et al., 2015; Gläscher et al., 2010; S. W. Lee et al., 2014; Smittenaar et al., 2013), while model-free prediction errors have been mainly localized to the (ventral) striatum (Beierholm et al., 2011; Gläscher et al., 2010; O'Doherty et al., 2015) or the putamen (Doll et al., 2015, but see also Daw et al., 2011; Sanfey & Chang, 2008). A potential causal role of the DLPFC in model-based decision-making was identified via direct manipulation of the DLPFC via TMS, which led to a reduction in model-based decision-making (Smittenaar et al., 2013).

In contrast, only a few studies have addressed the neural correlates of metacontrol concerning switching between decision-making strategies (S. W. Lee et al., 2014; O'Doherty et al., 2015). For example, O'Doherty et al. suggested that the arbitration between model-free and model-based systems was encoded by bilateral inferior lateral PFC, the right frontopolar cortex, and the rostral anterior cingulate cortex (O'Doherty et al., 2015). Meanwhile, Lee et al. found that the arbitration between habitual and goal-directed systems depended on activity in the bilateral lateral PFC (S. W. Lee et al., 2014). In addition, a study on adolescents found that the selective upregulation of cognitive control for trials with greater reward in contrast to trials with lesser reward was governed by frontostriatal connectivity (Insel et al., 2017). This could lead to a similar relationship in the context of stake-based metacontrol used in the current study. Taken together, findings from these studies suggest that DLPFC, in particular, may be implicated in both model-based decision-making and its metacontrol, however, presumably serving different respective functions.

In Chapter 2, I will explore model-based decision-making in childhood from the perspective of an adapted sequential decision-making task, where I find that it might appear much earlier than previously thought. In Chapter 3, I will investigate which EFs and individual differences in brain anatomy may be linked to and support model-based decision-making in childhood.

1.4 A rational or irrational discount: intertemporal decisions

In the first section of this introduction, I outlined the different schools of decision-making in terms of normative, descriptive, and prescriptive approaches and the type of choices that the field of judgment and decision-making coined as *decision-making under uncertainty* (Erez & Reyna, 2019). Many of the decisions that people make daily fall under this category, with some decisions having a substantial long-term impact on our life (e.g., choosing a field of study, deciding to move abroad, choosing whom to marry). Other decisions are on a much shorter term, but may still have a significant impact (e.g., conducting a dangerous over-taking maneuver on a busy road, having unprotected intercourse). As I stipulated previously, most choices require people to trade off costs and benefits at different points in time. Decisions that have consequences in multiple periods are referred to as intertemporal choices. Therefore, decisions about savings, work, effort, education, nutrition, exercise, avoiding climate change, and health care are all intertemporal choices (Chapman, 1996; Hamilton & Potenza, 2012; Kacelnik, 1997; Slavic et al., 1977).

Understanding how people's decisions are affected when faced with different temporal outcomes is essential and critical in understanding how society tends to tackle challenges and how our decisions may be influenced and manipulated. For example, environmental policies often require trading consequences with a differing time horizon: the immediate loss of gain resulting from banning forest cutting against the delayed, long-term loss caused by losing biodiversity and atmospheric activity of that forest. To make these types of decisions, people must, consciously or unconsciously, combine the magnitude of the immediate and future consequences and the time delay until each of them, which requires a time-discounting criterion.

The field of normative decision-making is concerned with defining optimal, rational decision-making, free from biases, and stipulated how people *ought* to make rational decisions (Slavic et al., 1977). Inspired by economics, scientists were interested in whether people adhered to these logical principles in financial decision-making. Normative theories posit that minimal discounting with time is a valid normative construct (Elliott, 2019; Kacelnik, 1997). A problem with normative approaches to temporal discounting is that it is not evident if there is only one way to be rational about discounting. Psychological research cannot replace the need for ethical, psychological, and economic input for socially meaningful decisions such as environmental or health care policies. In the case of a more personal argument, one's approach to saving is influenced by how one feels about postponing immediate gratification and external advice about the theoretical expectations of economic performance for the money set aside. Another factor is trust and perceived stability in one's personal life. Without trust in financial institutions, one may be less likely to put money aside in savings or investment accounts (Guillou et al., 2020, 2021). If there is instability in one’s personal life, for example, a family struggling to make ends meet, it may be preferred not to “risk” the uncertainty of time but rather to choose a guaranteed pay-off today.

In descriptive decision-making approaches, which are concerned with analyzing how people *do* make decisions based on empirical approaches, it quickly became apparent that people discounted rewards that took place further away in the future. Temporal discounting, or the decrease of utility or value of a reward over time, has been a consistent finding in decision-making research (Keidel et al., 2021; Kirby, 2009). The theory of discounted utility is the most widely used framework for analyzing intertemporal choices (Chapman, 1996). Frequently, hyperbolic or exponential curves are fitted to choice data, and the parameters that dictate the steepness of the downwards curve are used to determine a person’s individual tendency to discount a reward with time (Kacelnik, 1997). However, recent theories have suggested that fitting these curves may not provide a complete *descriptive* account of the cognitive influences on intertemporal decisions in humans (Bos & McClure, 2013).

Commonly, rewards that are available more immediately are preferred over rewards that may be larger but are delayed. This preference is known as temporal discounting, and this preference has been demonstrated in intertemporal choice studies where individuals are asked to choose between a smaller sum of money immediately (e.g., 10 dollars now) or a larger delayed sum (50 dollars in a week); the further away the reward, the less the reward is valued (Bos & McClure, 2013; Kirby, 2009). However, from a purely economic standpoint, the way we process delay should be constant, and we should not devalue a larger reward because it is only available further in the future[3]. The most interesting question, then, becomes what factors predict individual differences in tendencies to discount rewards more with time. Reduced sensitivity to temporal discounting, which is therefore preferred, may be supported by episodic memory (Bulley et al., 2016; Shohamy & Daw, 2015), working memory (Wesley & Bickel, 2014; Zhao et al., 2022), intelligence (Rustichini, 2015), and cognitive control (Figner et al., 2010; Steinbeis et al., 2012, 2016). On the other hand, individual variations in how steeply rewards in the future are discounted have been linked to psychopathology (Moutoussis et al., 2021; Story et al., 2014), problems with processing memory (Mellis et al., 2019; Wesley & Bickel, 2014), and behavioral disorders such as gambling addiction (Bickel et al., 2007, 2014). In children, developmental studies show that with age, children become increasingly more patient (Green et al., 1999; Prencipe et al., 2011), which translates to them discounting future rewards less steeply with time. However, the mechanisms underpinning this development, for example, whether improvements in memory or other EFs in childhood may be driving this move towards more patient decision-making, remains unclear. In Chapter 4, I will explore the associations between EFs and intertemporal decision making in childhood.

1.5 The social aspect in rational decision-making

In the previous section, I discussed the commonly observed effect with which individuals may discount rewards set in the future with progressive time and

[3] This is assumed rational in the absence of strong inflation.

how this can also impact social decisions, for example, regarding climate change and health care policies. Human beings are naturally social creatures, and many of our daily decisions also impact the people around us and, therefore, take place in a social dimension.

Our decisions can also affect our social lives and functioning, as daily we engage in the social exchange of goods and money (Henrich et al., 2005). These exchanges can pose a conflict of interest, where both parties aim to maximize their outcomes (Guth et al., 1982). For example, imagine a child at school engaging in a popular card trading game. The child may have doubles of a card, which they want to exchange for a desired card missing from their collection that another child possesses, the goal being to get the best deal possible. Therefore, the child needs to make an acceptable offer to the other child to meet their goal and, ideally, preserve a friendship so they can potentially trade again in the future if needed (Steinbeis et al., 2012). Cooperation thus involves a delicate balance of achieving one’s own goal, understanding the other person’s goal, and maintaining good standing for future interactions (Wang & Liu, 2022). Being known for generosity and sharing fairly with peers can help one gain social capital and develop successful ongoing and new reciprocal relationships (Bull & Rice, 1991). Instead, making more selfish decisions at the expense of the people around us, for example, stealing money or items, can make people more cautious or even refuse to interact with us in the future (Fehr & Gachter, 2000).

The human ability to intuitively assess fairness is a critical factor in our capacity to cooperate within a larger society of unrelated and often unfamiliar individuals. One of its key components is the ability to divide resources equally among members of a society (Rawls, 1971). How individuals respond to inequality in the distribution of resources is a useful objective benchmark of one's underlying sensitivity to fairness (McAuliffe et al., 2017). A consistent and cross-cultural observation is that adults dislike receiving less than others, and adults prefer to receive nothing than accept inequality (Dawes et al., 2007). Further, adults have been shown to punish the proposers of unfair distributions even at cost to themselves (Dawes et al., 2007; Henrich et al., 2005). When participants are put in a situation of control over the division of resources in the absence of sanctioning threat, dividing the resources equally has been considered a measure of altruism or unconditional fairness (Benenson et al., 2007; Edele et al., 2013; Hilbig et al., 2015). Pro-social decision-making has previously been linked to a better Theory of Mind (Li et al., 2017; Santamaría-García et al., 2018; Wang & Liu, 2022), empathy (Q. Guo & Feng, 2017; R. Guo & Wu, 2021; Herne et al., 2022; Zhang & Wang, 2019), and personality traits (Allgaier et al., 2020; Gummerum et al., 2010; Hilbig et al., 2015). Moreover, it has also been linked to better EFs, for example, in the context of cognitive control (Figner et al., 2010; Steinbeis, 2016; Steinbeis et al., 2012; Steinbeis & Over, 2017).

As previously mentioned, children increasingly make more pro-social decisions with age, which may be linked to an increase in inequality aversion (Blake & McAuliffe, 2011; Fehr et al., 2008; Gummerum et al., 2008). However, even if children understand a situation to be unfair and judge it so, this is not necessarily reflected in their subsequent behavior (Blake et al., 2014; Smith et al., 2013; Steinbeis et al., 2012) (however, also see (Paulus et al., 2018)). While studies of adults cannot differentiate between the processes acquired through society and those with deeper biological roots, studies of children can help distinguish between the foundational and malleable components of social decision-making (McAuliffe et al., 2017).

Interesting, there must be intrinsic motivation to engage in pro-social behavior as humans invest time, money, and effort in others even without the chance of repeated encounters (Steinbeis, 2016). Such altruism, which is defined as behaviors which incur a personal cost to benefit another in some way (Fehr & Fischbacher, 2003), already occurs early on in development, and in addition can be observed in other species, such as chimpanzees (Brosnan et al., 2010; Warneken & Tomasello, 2006). Altruistic behaviors can include helping, comforting, and sharing of resources (Schmidt & Sommerville, 2011; Svetlova et al., 2010; Warneken & Tomasello, 2006). As mentioned before, children have been found to become more pro-social with age. However, which psychological mechanisms may underlie pro-social sharing of resources remains elusive, especially in the middle childhood period. In Chapter 4, I will discuss age-related changes in social decision-making, and its relation to EFs in childhood.

1.6 The developmental lens: executive functions and bounded rationality in decision-making

The concept of bounded rationality argues that humans are rational creatures, within bounds. How far these bounds stretch is potentially dictated by our cognitive abilities, such as our ability to search through different options when considering which decision to make or our ability to access previous associations or memories when faced with a similar scenario (Sutton & Barto, 2018). In psychology, cognitive abilities that encompass our ability to expend attention and focus on a task at hand, or our ability to flexibly shift between different tasks, are often considered to be *executive functions* (EFs) (Diamond, 2013). EFs are broadly defined as functions in the realm of working memory, cognitive flexibility, and cognitive control, or the ability to inhibit prepotent impulses (Diamond, 2013). Thus, an individual with a large working memory span and manipulation ability, good ability to flexibly switch between alternatives and contexts, and cognitive control over their actions, may therefore make more rational decisions.

Human decision-making can either be split into quicker, cognitively cheap, and habitual decisions or slower, cognitively expensive, and goal-directed decisions (Daw et al. 2011). While habitual decisions make up most of our daily actions (e.g., commuting home via the usual route, putting on the same t-shirt, making a morning coffee the same way), when faced with a new or difficult scenario, our goal-directed decision-making is prone to taking over (Dezfouli & Balleine, 2013; Dickinson, 1985). Using goal-directed decision-making allows us to approach a problem consciously, consider alternatives, and pick the right option (Gillan et al., 2015; Wan Lee et al., 2014). Because this reflects a sophisticated way of thinking and approaching decisions, engaging in effective goal-directed decision-making is often thought to be a late-developing skill that only becomes available after brains have matured (Decker et al., 2016; Nussenbaum et al., 2020; Nussenbaum & Hartley, 2019; Palminteri et al., 2016). Thus, another bound that may be placed on our rationality is the constraint of development.

Research in developmental studies, especially studies that span the lifespan from childhood to adulthood, is often concerned with tracking the development of the building blocks of reason and rational decision-making. A common finding has been that “higher” EFs, such as complex cognitive abilities, only become available due to the ongoing maturation of brain regions and connections, as childhood is a period of greater plasticity (Buss & Spencer, 2014; Fiske & Holmboe, 2019; Satterthwaite et al., 2013). Thus, if humans only reach their potential with developmental maturity, this means that rational decision-making should be developmentally incomplete.

For example, as previously discussed, research into changes in temporal discounting across the lifespan observed that in developmental samples, progressive age is linked to less steep temporal discounting (Green et al., 1999; Prencipe et al., 2011; Steinbeis et al., 2016). In addition, when we consider pro-social decision-making, as discussed in the previous section, developmental research has found that children’s decisions become progressively more pro-social with age (Bauer et al., 2014; Chajes et al., 2022; Fehr et al., 2008; McAuliffe et al., 2017). In a mirror view of this, when we consider the potential factors that may “bound” our rationality, such as working memory, cognitive flexibility, and cognitive control, we again observe a consistent increase with age (Chevalier, 2015; Davidson et al., 2006; Domenech & Koechlin, 2015; Ganesan & Steinbeis, 2022; Garon et al., 2014; Prencipe et al., 2011; Satterthwaite et al., 2013; Wiebe & Karbach, 2017). Potentially, our ability to expand our bounds may, therefore, underlie our ability to engage in rational decision-making.

While economic decision-making provides insight into how we might allocate resources at the moment or how long we can be persuaded to wait for a reward, it does not necessarily illuminate how we might learn from our past decisions. The field of reinforcement learning has been trying to formalize different ways of decision-making for nearly a century, from formalized models for animal decision-making to current research on complex decision-making strategies. A distinction between habitual and reflexive versus deliberate decision-making processes has been proposed in previous literature. This distinction ties in with the studies conducted in social and intertemporal decision-making, as it might require the ability to deliberate and correctly assess the outcomes of an action to see that the larger delayed reward has a higher payoff in the end.

Throughout the experimental chapters, I will explore social and intertemporal decision making and reinforcement learning from a developmental focus. I will discuss research on the potential underpinnings of more pro-social and intertemporal decision making, as well as a higher degree of model-based decision making. In Chapter 2, I will discuss the presence of model-based decision making in childhood, and that this may emerge much earlier than previously thought. In Chapter 3, I will review the potential underpinnings of individual differences in model-based decision making in childhood. In Chapter 4, I will discuss social and intertemporal decision making from a developmental and prescriptive lens.

1.7 The potential plasticity of social and intertemporal decision-making

Throughout this introduction, I discussed the history of decision-making, and its shift from the assumption of human beings as rational decision-makers in normative theories, to describing human decision-making, including all its inconsistencies and biases in descriptive theories. The third branch of decision-making research concerns prescribing how human beings should make decisions, the so-called *prescriptive* branch of decision-making research.

Based on the types of decision-making I have discussed so far, the decisions one *ought* to make seem straightforward. For example, one should engage in goal-directed decision-making when it is necessary to solve a complex problem. One should generally be pro-social and fair to maintain healthy and reciprocal social networks with access to fairly distributed resources. Furthermore, one should generally aim not to be swayed to strongly discount time but focus on future outcomes almost as readily as immediate outcomes so that one can save effectively and lead a successful life by planning toward future goals. However, as I reviewed early on, humans are not rational decision-makers that normative theories and philosophies can capture. Instead, human decision-making is inconsistent, messy, and often irrational.

When the concept of bounded rationality was introduced, it was a way to explain this absence of rationality in human decision-making. If we can only maintain so much information in our minds at a time, we can only be rational to a certain extent. However, with the introduction of this theory, Miller also hinted at individual differences in this regard in the title “The magical number seven, plus or minus two: Some limits on our capacity for processing information” (G.

A. Miller, 1956). If we could extend the limits, would we, in turn, also see increases in our rational decisions?

These limits may be dictated by our EFs, as they are critical to supporting goal-directed behavior (Diamond, 2013). Additionally, childhood executive functioning has been shown to predict various social, academic, and mental health outcomes later in life (Blair & Razza, 2007; Clark et al., 2010; Moffitt et al., 2011). As we mentioned before, EF ability undergoes protracted development from childhood into early adulthood (Davidson et al., 2006; Garon et al., 2014; Wiebe & Karbach, 2017), likely supported by accompanying changes in frontoparietal and frontostriatal anatomy and connectivity (Buss & Spencer, 2018; Fiske & Holmboe, 2019; Insel et al., 2017). Given their strong links to real-life outcomes and prolonged plasticity throughout development, EFs have been primary targets for brain training interventions (Diamond & Lee, 2011; Wass et al., 2012). While several training paradigms have had successes in improving the trained domain, for example, several studies have found that training working memory capacity does lead to long-term increases in working memory span (Schmiedek et al., 2010), the ultimate goal of these training paradigms is to test whether improvement in one cognitive ability also leads to improvements in a loosely related domain (Smid et al., 2020; Wass et al., 2012; Wilkinson et al., 2019). However, training studies that report robust effects on loosely related domains are few and far between, and the current consensus is that improvements in a correlated function may not effectively translate further than the trained subject (Gobet & Sala, 2022; Sala & Gobet, 2016, 2017, 2019). While many studies have investigated how EF training might impact other cognitive abilities or academic outcomes, not many studies have investigated whether improvements in EFs via a training paradigm may lead to changes in social or intertemporal decision-making (but see (Kable et al., 2017; Steinbeis & Over, 2017; Zhao et al., 2022)). Intervention studies in childhood may be particularly interesting because this developmental period is marked by substantial changes in EFs and decision-making (Wass et al., 2012). Only a handful of studies have investigated how training EFs may impact decision-making in childhood specifically (Steinbeis & Over, 2017; Zhao et al., 2022).[4] In the final experimental chapter of this thesis, I will investigate if training cognitive control led to increases in pro-social decision-making and less steep temporal discounting in intertemporal decisions.

1.8 Summary

There are three main branches on approaching decision-making research: *normative*, *descriptive*, or *prescriptive* theories. Normative theories concern itself with how humans *ought* to decide, while descriptive research investigate how humans *actually* decide. Applying these two concepts, prescriptive research seeks to make humans more normative decision-makers. Descriptive research has shown that humans do not adhere to normative theories in their decision-making. For example, on a daily basis humans make decisions that impact their health and social wellbeing. Humans have been shown to strongly discount consequences of positive outcomes that are set further away in time.

A pervasive explanation of why humans do not seem to be rational, normative decision-makers, is offered via the theory of bounded rationality

[4] Steinbeis & Over, 2017 used a behavioral control priming paradigm, rather than the actual training of the ability.

which proposes that humans are rational within the limit of their cognitive abilities. In addition, the concept of dual-systems theory has led to a pervasive theory of human decision making: humans rely on both a quick, intuitive habitual system and a slower, deliberate goal-directed system to make decisions. Goal-directed decision making reflects a conscious and deliberate approach to decision making, which more closely aligns to the normative theories. The field of reinforcement learning uses computational models to quantify decision-making behavior and is able to capture how much of each system one relies on for their decisions. By quantifying the individual differences in habitual and goal-directed decision making, these differences can be linked to other cognitive abilities.

Studies investigating the development of decision-making in humans report that children make decisions that show a lack of pro-sociality and inability to effectively consider intertemporal outcomes for decisions when compared to adults. Childhood is marked by substantial changes and increases in decision making behavior and cognitive abilities, mirrored by structural and connectivity related changes in brain development. From a prescriptive lens, if cognitive abilities are a limiting factor in rational, normative decision making, enhancing cognitive abilities will potentially translate to improvements in rational decision-making. With childhood being a critical developmental period marked by greater plasticity, or the capacity to undergo substantial changes, it may therefore be a particularly effective period for intervention studies to enhance cognitive abilities and to impact decision making.

This thesis aimed to shed light on learning behavior in the context of decision-making in childhood, using an interdisciplinary computational approach via reinforcement learning models. Further, it sought to investigate whether goal-directed decision-making was linked to individual differences in EFs and brain anatomy. In addition, it sought to revisit the relationships between EFs and social and intertemporal decision-making, how these relationships change across childhood, and whether training EFs may lead to improvements in intertemporal decisions and pro-social decision-making. For all three experimental chapters, accompanying data and code for the analyses can be found on my Github (https://github.com/ClaireSmid).

Chapter 2. Model-based decision-making in childhood

Part of Chapter 2 was published in a research paper in Developmental Science:

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

2.1 Abstract

Human decision-making is underpinned by distinct systems that differ in flexibility and their associated cognitive cost. A widely accepted dichotomy distinguishes between cheap but rigid model-free and flexible but costly model-based systems. 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, I demonstrate that children aged 5 to 11 years (N = 85), including the youngest children, displayed multiple indicators of model-based decision-making and 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. My results demonstrate that children can engage in highly sophisticated and costly decision-making strategies throughout childhood. However, the flexible

arbitration between decision-making strategies might be a critically late-developing component in human development.

2.2 Introduction

To navigate our world successfully, we need to learn which of our actions lead to desirable outcomes. It is commonly theorized that human reward-related learning is guided by at least two decision-making systems competing for control (Daw et al., 2005; Gläscher et al., 2010). One is a goal-directed and computationally 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, enabling the repetition of previously reinforced actions (Dickinson et al., 2002). The field of reinforcement learning provides a practical 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 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 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 environmental changes. 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, 2019). 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 a more demanding but more accurate model-based system. While a wealth of studies show that adults use both systems when making decisions, little is known about how these systems come to contribute to decision-making during human development.

Children can make simple value-based decisions from a young age by learning which actions lead to positive and negative outcomes. For example, even young infants have been shown to link actions and reward through gaze following (Ishikawa et al., 2020), to learn the underlying hierarchical structure of a sequential decision-making task (Werchan & Amso, 2021), and to understand goal-directed movement (Southgate et al., 2014). 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). In addition, the ability to control reflexive responses to stimuli, an executive function named *inhibition*, is present from a young age and continues to improve and develop further through childhood (Davidson et al., 2006). While these studies show that children can learn the relationship between their actions and subsequent reward, it is unclear whether children rely on model-free action-reward contingencies or 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 to 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., 2015; 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. However, 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 plan effectively through the task structure (Kool et al., 2016). Indeed, recent variations of the traditional two-step task that simplified the transitional structure, which do allow a model-based system 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 to 12-year-old children cannot 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, I 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, I 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 model-based decision-making when greater rewards could be won (Bolenz et al., 2019; Kool et al., 2017; Patzelt et al., 2019). However, whether or not children engage in effective and flexible metacontrol over distinct decision-making systems remains unclear. Therefore, in addition to investigating whether children of this age range could engage in model-based decision-making, I 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, I used 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, I address two questions; first, whether children aged 5 to 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, the current results suggest that pre-adolescent children can engage in model-based decision-making, which I 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.

2.3 Materials and Methods

2.3.1 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. I 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). Eppinger et al. (2013) found large age-related effects in model-based decision-making in an adult sample with 60 participants. The information entered into the power calculation was an α of .05, a power of 90%, and the effect size found by Eppinger et al. ($t < 4.04$, $p < .001$, $\eta 2 = 0.20$) (Eppinger et al., 2013; Faul et al., 2007). Additionally, Decker et al. (2016) found an age-related effect of model-based decision-making across their sample with 59 participants (children, adolescents, and adults), with a medium effect size, ($X2(1) = 26.00$, $p < .001$, effect-size estimate = 0.27, se = 0.05) (Decker et al., 2016). Based on this, I determined that with a sample size of at least 60 children, I would achieve more than 90% power to detect a true age-related effect of comparable size.

A total of 114 children were tested in a classroom locally in a school in Greater London. Every experimenter tested two children simultaneously, and five experimenters conducted testing in the classroom each day. Children were tested in between classes with the help of teaching staff. Due to time constraints, some participants could not complete the entire task. Therefore, I 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 the task was 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 comprised 85 participants (37 girls (44%) and 48 boys). The mean age of children was 8.2 years (*SD* = 1.6), ranging from 5.0 to 11.4 years, see Table 1. 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.

Table 1. Ages and gender for the developmental and adult sample

	Total N	Subset N	Mean age (SD)	Gender (F)
Children	85		8.2 (1.6)	43.5%
5-year-olds		7	5.6 (0.3)	42.9%
6-year-olds		18	6.4 (0.3)	44.4%
7-year-olds		15	7.6 (0.2)	46.7%
8-year-olds		16	8.5 (0.3)	50.0%
9-year-olds		16	9.5 (0.3)	37.5%
10-year-olds		11	10.6 (0.3)	18.2%
11-year-olds		2	11.2 (0.2)	50.0%
Adults	24		25.2 (4.7)	45.8%

2.3.2 Sequential decision-making task

2.3.2.1 Task design 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 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 who gave the participants treasure when they visited their planet. To be manageable for the younger children in the sample, the current task consisted of 140 trials (compared to 201 trials in Kool et al. 2017). Therefore, I 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 the sample. For these results, please see 2.6.3 Parameter recovery.

Trials consisted of two stages. In the first stage, participants were randomly presented with one of two possible pairs of spaceships displayed on an earth-like planet background (see Figure 8a). Each spaceship appeared on the left and right sides with equal probability. 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 (Figure 8a). The choice between the two spaceships had to be made by pressing one of two keyboard keys (i.e., “F” or “K”) within a time window of 2 seconds. After a spaceship was selected, it was outlined by a border for the remainder of these 2 seconds, meaning trials could not be progressed through faster but had a fixed duration.

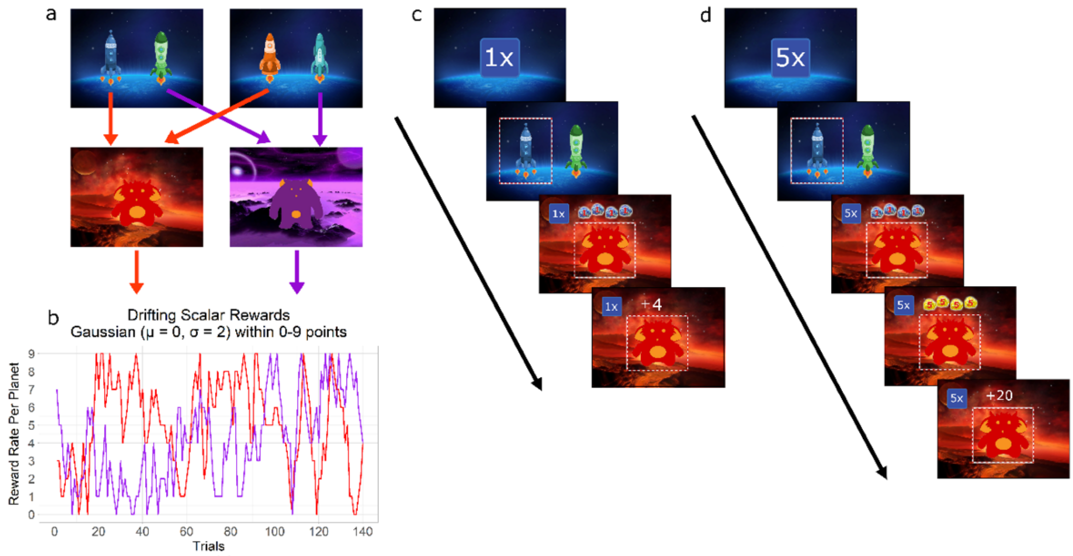


Figure 8. Sequential decision-making task design.

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 0-9 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.

In the second stage, participants transitioned to either a red or a purple planet, as determined by their choice in the first stage. Note that both first-stage states offered the possibility to visit either planet at the second stage, with one spaceship always going to the red planet and the other to the purple planet.

When arriving at the planet, they saw an alien and had to press the spacebar within 2 seconds to collect “space treasure” from them (e.g., see Figure 8c and Figure 8d). The reward distributions between the two planets were initialized within a range of 0 to 4 points for one planet and 5 to 9 points for the other. Afterward, the reward distributions varied according to a Gaussian random walk (standard deviation = 2) with reflecting bounds at 0 and 9 for the rest of the experiment. They were told that the aliens slowly moved between the bad and good parts of their cave to make the participants aware of this fact. A new set of randomly drifting reward distributions was generated for each participant (Figure 8b). Such drifting reward rates have been shown to promote learning and continuous monitoring of rewards won at each planet, allowing a model-based system to capitalize on faster changes in rewards compared to the traditional two-step task (Kool et al., 2016).

Importantly, the spaceships in the first states were practically equivalent. One spaceship in a pair always led to the red planet, and the other spaceship always led to the purple planet. Because of this equivalence, model-free and model-based contributions to decisions can be dissociated since only the model-based system generalizes across the equivalent starting states. 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. Thus, 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.

2.3.2.2 Stakes manipulation

I employed low and high-stake trials to test whether the participants arbitrated between employing model-free and model-based systems depending on the rewards available. During the trials, participants received rewards in the form of pieces of blue space treasure. At the start of each trial, participants were randomly presented with one of two “treasure amplifiers” for 2 seconds. These treasure amplifiers indicated whether the trial was a low-stake (the amplifier showed “1x”, and the number of points they received was the same as pieces of treasure shown) or a high-stake trial (the amplifier showed “5x”, meaning that it was worth five times more points, see Figure 8c and Figure 8d). On a low-stake trial, the pieces of treasure won directly translated to the number of points won on that trial, e.g., four pieces of blue treasure would have a value of four points (Figure 8c).

In contrast, during a high-stake trial, rewards were multiplied by five, e.g., 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 (Figure 8d). High- and low-stake trials were at an approximate 50/50 ratio and occurred randomly.

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, I 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.

2.3.2.3 Instruction phase

All participants completed an identical instruction phase, which took approximately 20 minutes, and the main task itself took approximately 25 minutes to complete. The instruction phase was identical for children and adults. No rewards were gained during the instruction phase, and practice trials were not used for further analysis.

During the instruction phase, participants 1) observed a demonstration of the drifting reward rates, which showed how fast the amount of treasure could change over trials. Participants observed two drifting reward distributions of 5 examples per planet. This was fixed identically for all participants. Participants had to verbally report to the experimenter after each completed demonstration whether the treasure increased or decreased over the examples and were corrected if wrong and given feedback; 2) then, they completed a training stage to practice the transition structure. Here, a criterion of four correct consecutive transitions to the red and purple planet, respectively, were required to pass for all participants. After ten tries without successfully passing the requirement, participants were reminded of which planet they needed to travel to. Therefore, the task would only continue after participants had learned to successfully deterministically transition to each planet; 3) to familiarize participants with the trial sequence; in the third section, they completed six practice trials without stakes where they traveled to a planet of their choice and collected treasure using the transition structure. Unlike trials in the main task, no stake cue was presented, and the trials did not time out, whereas the trials in the main task had a 2-second response window. Additionally, treasures won were not kept; 4) the last phase was a stake instruction where participants saw the stake cue, then a picture of one of the planets, randomly chosen, and then the trial animation associated with the respective stake together with the stake cue in the corner (“1x” or “5x”; see Figure 8c and Figure 8d). Treasure would be displayed above the alien with a question mark. Participants then had to verbally state how many points they would receive for the treasure shown, e.g., if they saw four pieces of treasure, during low-stake trials, the correct answer would be 4, and during high-stake trials, this would be 20. Participants stated their answers to the experimenter, who provided feedback and corrected them if wrong, and then again explained the stake condition. No full trials were played, and no points could be earned.

At the start of the main task, the drifting reward rates for each planet were reset, and any learned associations between rewards and planets during the instruction phase were irrelevant for the main task. As a result, none of the trials from the task preparation phase were included in the computational models.

After the task preparation phase, all participants (children and adults) were asked to report on the transition structure. Participants did this by noting on a colored print-out sheet which spaceships they thought traveled to the red and the blue planets. Spaceships were presented next to each other and not displayed in pairs, as was the case on the computerized task. The transition structure differed between participants (e.g., which spaceships traveled to which planets), and whether the participants' answers were correct was therefore only assessed after testing and had no influence on the further testing procedure. However, the practice phase had a criterion of four consecutive accurate transitions to both planets to pass; this ensured that all participants learned the transition structure. The transition mappings of spaceships to planets were random for all children. For the adults, one fixed mapping of spaceships to transitions was used.

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 gathered 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, ($t(66) = .43$, $p = .670$, 95% CIs [-.17, -.26]), suggesting that the majority of the children learned the deterministic structure of the task.

2.3.3 Statistical analysis and corrections

All statistical tests were conducted in R. For general effect sizes, I report 95% confidence intervals and Cohen's d, and for regression results, I 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 samples 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 contain 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 evaluated using the PPCOR package (S. 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, 2018). In computational models, *priors* can be used, which are values used to initialize the parameters of a model. Using priors helps with the accuracy of model-fitting. If priors are kept “vague”, they do not influence the parameter solution strongly and only have a minimal effect on parameter solutions. Therefore, I 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). I used Beta(2,2) priors for all model parameters bounded between 0 and 1 (learning rate (a), eligibility trace (λ), and the mixing weight(s) *w*), and a Gamma(3,0.2) prior for the inverse Softmax temperature (b), and for the two choice stickiness parameters (π and r) I used Normal(0,1) priors (Bolenz et al., 2019).

The model-fitting procedure I used to acquire the parameter solutions has the potential to introduce noise. To avoid this, I used model-free simulations to create a baseline to which I could compare the children (see 2.3.4 Model-free simulation procedure). For more details on the dual-systems reinforcement-learning model see 2.6.1 Dual-reinforcement learning model, for the model-fitting procedure see 2.6.2 Model-fitting procedure, for model performance and comparisons see 2.6.4 K-fold cross-validation, 2.6.5 Bayesian model comparison and 2.6.6 Qualitative model validity.

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, I used the ggeffects package (Lüdecke, 2018).

2.3.4 Model-free simulation procedure

An essential aim of this study was to investigate whether children showed influences of model-based behavior. However, since the model-based weighting parameter is bounded between 0 and 1, estimates of this parameter will always be larger (or equal) to zero, which means that noise in either the model-fitting procedure or in behavioral performance can only push this parameter over the lower bound, and not under. Therefore, I needed a more meaningful baseline value for model-free decision-making. I established this baseline as the average estimated weighting parameters of simulated data from agents that were completely model-free but otherwise matched with the sample.

We simulated 500 iterations of 85 synthetic model-free agents. I generated these agents the following way. First, I fitted a 5-parameter model, a reinforcement learning model with w hardcoded to 0, to the decision data from the children. Next, I used the parameter solutions from the other parameters in this model (inverse temperature, learning rate, etc.), with added noise, and again set w to 0 to fit a generative version of the 5-parameter model to simulate behavior that would be completely model-free. Next, I fitted the 6-parameter model that included a free w-weighting parameter for model-based decision-making over the whole task to this simulated model-free behavior. I used the w-value that came from the 6-parameter model as the model-free baseline. These simulations provided us with a true null baseline value for model-free decision-making as determined by the reinforcement learning model that can be used to compare data from actual participants meaningfully.

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

2.4 Results

2.4.1 Children perform above chance level and are not random

To assess whether children were sufficiently engaged with and capable of doing the task, I 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 8b). This corrected reward rate tracks task performance against chance level (0). Scores lower than 0 indicate performance worse than chance, and scores higher than 0 indicate better than chance performance.

There was no difference in mean reward rate (the drifting reward rates), or, the amount of reward that could be won, between the adults (M = .50, SD = .04), and the children (M = .49, SD = .05), (t(1, 107) = -.30, d = -.007, p = .768, 95% CIs [-.02, .02], Figure 9a).

The mean corrected reward for children was significantly higher than chance (t(84) = 3.20, d = .35, p = .002, 95% CIs [.003, .013]). Performance was also significantly correlated with age (r = .32, p = .003, 95% CIs [.12, .50], Figure 9b). This suggests that the children were meaningfully performing the task and that performance improved throughout childhood.

To compare reaction times between children and adults, I log-transformed the reaction times. Between children (M = 0.58, SD = 0.23) and adults (M = 0.56, SD = 0.17), there was no significant difference in reaction time, (t(41.08) = 0.89, d = 0.19, p = .377, 95% CIs [-0.03, 0.07], Figure 9c), suggesting that children and adults could both successfully complete the task, and children were not at ceiling and therefore not too rushed to make their response. The two groups were matched on gender (X^2(1, $N = 109$) = .00, φ = .019, p = 1.000). For children, there was no significant correlation between log-transformed average reaction time and age, (r = .18, p = .098, 95% Cis [-0.03, 0.38]), Figure 9d).

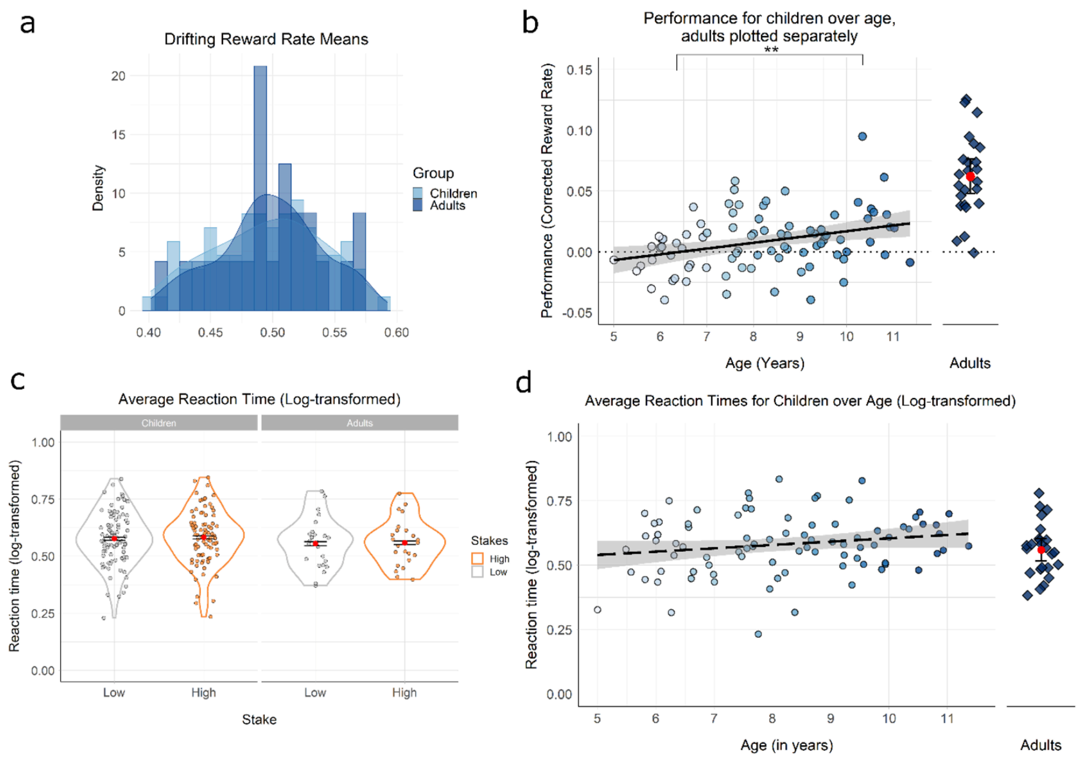


Figure 9. Performance and reaction time measures for children and adults.

a) Mean reward drift rate for all children and adults, plotted as density curves overlaid on histograms per group, (b) Performance metric for children plotted over ages with adults plotted separately, (c) Log-transformed mean reaction times per stake for children and adults, (d) Log-transformed average reaction times for children over age with adults plotted separately. Error bars depict 95% confidence intervals, and shaded areas around regression lines indicate the standard error of the mean. On graph c, distribution is shown as violin plots.

2.4.2 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 successfully engaging a model-free or a model-based system. Thus, I investigated whether children 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, I will name this parameter “model-based contribution” throughout.

For both children and adults, I conducted a formal model comparison where I 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). I compared the models using k-fold cross-validation, Bayesian model selection, delta AICs, and parameter recoverability in two separate parameter recovery analyses and a qualitative model assessment. From this comparison, the 6-parameter stake-agnostic dual-systems reinforcement learning model was the winning model overall. I fit the 6-parameter model to the data to assess model-based decision-making agnostic of stakes, and I use the 7-parameter model to evaluate metacontrol. For the full computational model, model comparisons, and parameter recovery analyses, see 2.6 Supplementary Materials.

First, I investigated whether children displayed model-based decision-making on the task over all combined trials. Children had an average model-based contribution of 0.52 (SD = .17) and given that this value is significantly larger than 0, ($t(84) = 27.40$, $d = 2.97$, $p < .001$ 95% CIs [.48, .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.

I created model-free simulations based on the children's data to resolve this. This resulted in a mean model-based contribution parameter of 0.28 (SD = .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.

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, ($t(84.22) = 12.47$, $d = 3.49$ $p < .001$, 95% CIs [.20, .27]).

Additionally, I investigated whether children's degree of model-based decision-making increased with age. I found that there was a positive relationship between the degree of model-based decision-making and age ($r = .22$, $p = .042$) (Figure 10a).

Furthermore, I investigated whether the youngest children also showed significant model-based decision-making. I 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, (Figure 10b), (5-year-olds: $N = 7$, $t(6.00) = 4.28$, $q = .005$, $d = 10.36$, 6-year-olds: $N = 18$, $t(17.01) = 6.53$, $q < .001$, $d = 7.32$, 7-year-olds: $N = 15$, $t(14.00) = 5.21$, $q < .001$, $d = 7.11$, 8-year-olds: $N = 15$, $t(14.00) = 3.95$, $q = .002$, $d = 5.41$, 9-year-olds: $N = 17$, $t(16.00) = 4.47$, $q = .001$, $d = 5.62$, 10 ($N = 11$) and 11-year-olds ($N = 2$): $t(12.00) = 8.65$, $q < .001$, $d = 13.39$).

I also investigated whether a linear model with a quadratic age term fitted the change of model-based decision-making better (e.g., (w ~ Age + Age^2)). When comparing a linear model with only age to a model with an additional quadratic age term, there was no significant difference between the models ($F(1,83) = 3.76$, $p = .056$). Thus, I interpret the changes in model-based decision-making over age as mainly linear.

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, I 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 ($r = .51$, $p < .001$), showing that a higher degree of model-based decision-making was significantly related to better performance. This effect remained significant after controlling for age ($r = .37$. $p < .001$).

Lastly, I inspected potential gender-based differences in model-based decision-making. First, I looked at whether there were overall differences in gender for the children. There were no overall differences between males and females in the degree of model-based decision-making ($F(1,83) = 0.01$, $p = 0.923$). When I ran a two-way ANOVA with Gender and age as predictors, only age was a significant predictor ($F(1,81) = 4.24$, $p = .042$), and there was no main effect of Gender ($F(1,81) = 0.01$, $p = .921$), nor an interaction between Gender and Age ($F(1,81) = 0.83$, $p = .365$).

For the adults, there was no overall effect of gender on model-based decision-making ($F(1,22) = 0.07$, $p = 0.800$). In a two-way ANOVA with gender and age as predictors, neither age nor gender were significant (Gender: $F(1,20) = 0.07$, $p = 0.800$: Age: $F(1,20) = 2.36$, $p = 0.140$), and there was no interaction, ($F(1,20) = 0.05$, $p = 0.826$).

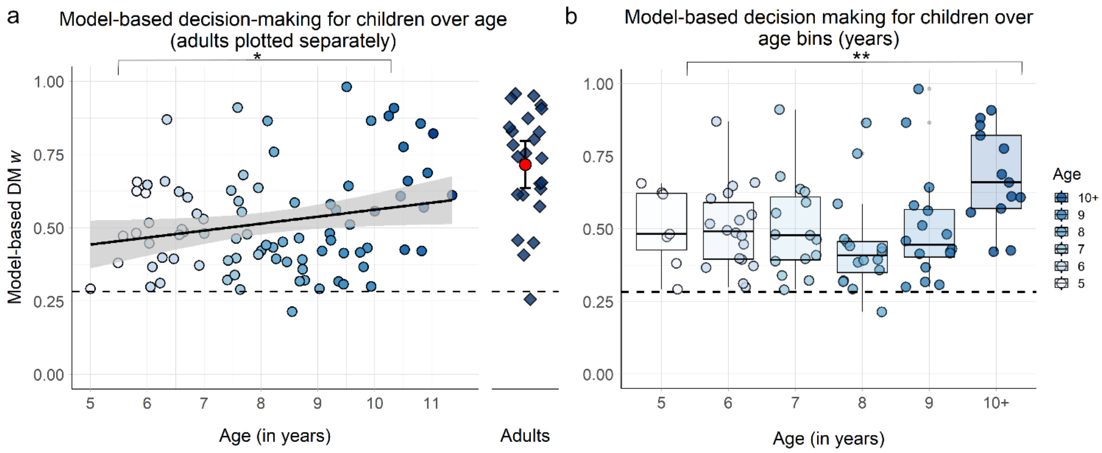


Figure 10. Model-based decision-making over age for children with the simulated model-free baseline.

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, I combined these children with the 10-year-olds (10+). The dashed line represents the simulated model-free baseline. Asterisks indicate significance level, *p<.05; **p < .01; ***p<.001. For panel b, significance indicates the highest *q*-value of each binned year of age against the model-free simulations.

2.4.3 Other computational parameters and age

We also looked at any other potential correlations with age for the children for the other model parameters for the 6-parameter model. The best-fitting parameter values are reported in Table 2.

Table 2. Best-fitting parameter values for the dual-systems reinforcement learning model without stakes (6 parameters) for children and adults

Groups and Predictors	β Inverse temperature	α learning rate	λ eligibility trace	π choice stickiness	ρ key stickiness	*w* model-based weight
Parameter bounds	[0,20]	[0,1]	[0,1]	[-20,20]	[-20,20]	[0,1]
Children						
25th percentile	0.47	0.31	0.45	-0.17	-0.74	0.39
Median	0.59	0.50	0.48	0.29	-0.30	0.48
75th percentile	0.79	0.66	0.54	0.62	0.41	0.62
Adults						
25th percentile	0.59	0.72	0.49	0.08	-0.19	0.61
Median	1.06	0.81	0.55	0.24	-0.06	0.75
75th percentile	2.15	0.89	0.58	0.78	0.07	0.85

None of the other parameters were correlated with age, (inverse temperature: r = .005, p = .965, 95% CIs [-.21, .22], Figure S2a; learning rate: r = .12, p = .270, 95% CIs [-.09, .33], Figure S2b; eligibility trace: r = .05, p = .655, 95% CIs [-.17, .26], Figure S2c; choice stickiness: r = .01, p = .925, 95% CIs [-.20, .22] Figure S2d; key stickiness: r = -.18, p = .101, 95% CIs [-.38, .04], Figure S2e).

Regarding correlations with performance for children, inverse temperature (r = .37, p < .001, 95% CIs [.18, .54]), learning rate, (r = .33, p = .002, 95% CIs [.12, .50]) and choice stickiness (r = .28, p = .009, 95% CIs [.07, .47]) were significantly correlated to corrected reward rate, while eligibility trace was marginally significant (r = .20, p = .064, 95% CIs [-.01, .40]). There were no significant relationships with performance for key stickiness: r = .01, p = .918, 95% CIs [-.20, .22]).

For the adults, only model-based decision-making during high stakes was significantly negatively correlated to age (r = -0.45, p = .029), showing that older adults had less model-based decision-making for high stake trials. None of the other parameters were significantly correlated to age (inverse temperature: $r = -0.12$, $p = .579$; learning rate: $r = -0.03$, $p = .889$; eligibility trace: $r = -0.09$, $p = .668$; model-based decision-making low stakes: $r = -0.07$, $p = .757$; choice stickiness: $r = -.14$, $p = .522$; key stickiness: $r = 0.03$, $p = .884$).

A high learning rate is optimal for this task since rewards continuously change, reflecting how much value participants place on recent information. In this task, the most recent information is the most valuable since that is the best way to stay updated on both planets' drifting reward rate distributions. Meanwhile, the inverse temperature signals how much exploitation and exploration participants employ. The eligibility trace reflects how much value is being placed on previously attained rewards on the task, or the reinforcement learning history, which is not highly important in this task. For the stickiness parameters, these would ideally be close to 0 to indicate an absence of bias in choices and keys.

We investigated which parameters differed significantly between the children and adults. Inverse temperature ($t(24.28) = 3.90$, $p < .001$), learning rate ($t(45.72) = 5.74$, $p < .001$), eligibility trace ($t(33.51) = 3.47$, $p = .019$), model-based decision-making during low-stake ($t(30.16) = 3.36$, $p = 0.025$) and high-stake trials ($t(30.00) = 4.35$, $p < .001$) were significantly higher for the adults than for the children. There were no significant differences in choice stickiness ($t(33.70) = 1.67$, $p = .104$) and response stickiness ($t(99.65) = 0.62$, $p = .535$).

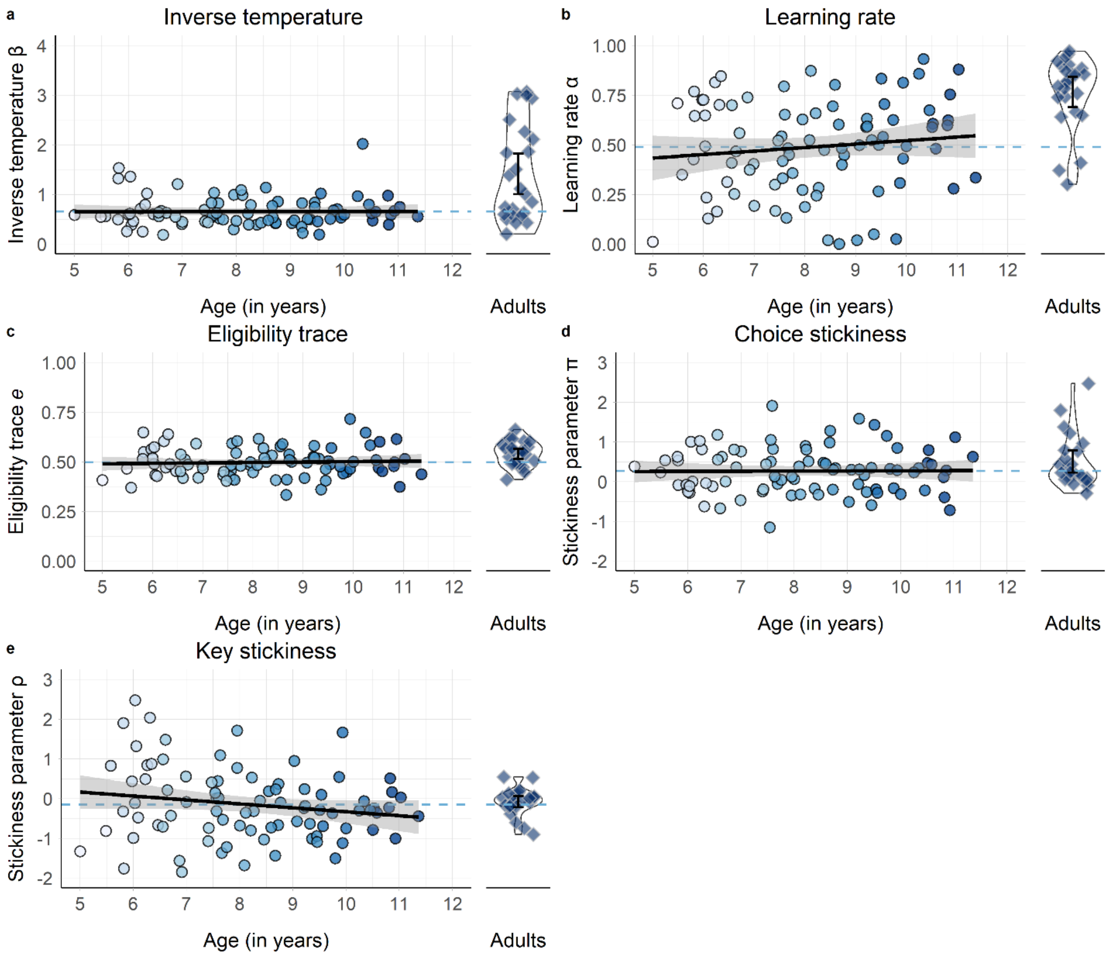


Figure 11. Other computational parameters over age for children and adults.

(a) inverse temperature, (b) learning rate, (c) eligibility trace, (d) choice stickiness, and (e) key stickiness. For all graphs, the shaded areas represent the standard error of the mean. The dashed lines in light blue represent the mean parameter values of the children.2.4.3 Metacontrol of decision-making for children and adults

2.4.3 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 (Figure 8c and Figure 8d). Any reward obtained on the trial is multiplied by five during high-stake trials, while on low-stake trials, the reward is multiplied by one and therefore does not change in value. The current task entailed changes to a previously used task with adults (Kool et al., 2016, 2017) (see 5.2 On the current and previous contrasting findings of model-based decision-making in childhood for details) 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, changes to task design (Figure 8c and Figure 8d)). I, therefore, first tested whether I could replicate a stakes effect in an in-person adult sample. To investigate this, I fitted adult data to a reinforcement-learning model that included a model-based contribution parameter that differed for each stake condition (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. I conducted *k*-fold cross-validation to investigate whether both models could reliably predict choices made by 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, 2.6.4 K-fold cross-validation). The best-fitting parameter values for the 7-parameter model for children and adults are represented in Table 3.

Table 3. Best-fitting parameter estimates for the dual-systems reinforcement learning model with stakes (7 parameters) for children and adults.

Groups and Predictors	β Inverse temperature	α learning rate	λ eligibility trace	π choice stickiness	ρ key stickiness	w_{low} model-based low	w_{high} model-based high
Parameter bounds	[0,20]	[0,1]	[0,1]	[-20,20]	[-20,20]	[0,1]	[0,1]
Children							
25th percentile	0.19	0.34	0.49	0.08	-0.19	0.44	0.43
Median	0.59	0.50	0.55	0.24	-0.06	0.51	0.50
75th percentile	0.80	0.66	0.58	0.78	0.07	0.57	0.60
Adults							
25th percentile	0.60	0.66	0.48	0.10	-0.18	0.50	0.63
Median	1.05	0.83	0.54	0.26	-0.04	0.57	0.74
75th percentile	2.15	0.88	0.62	0.81	0.08	0.72	0.84

Adults showed a higher degree of model-based decision-making during high-stake trials (M = .71, SD = 0.19), compared to low-stake trials (M = .61, SD = 0.18; $t(23) = 2.10$, $p = .047$, $d = .43$, 95% CIs [.001 .185]) (Figure 12a). This replicates previous findings of a stake effect on adult model-based decision-making (Bolenz et al., 2019; Kool et al., 2017; Patzelt et al., 2019).

Next, I 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, I 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, I found no significant difference in model-based decision-making between the low-stake (M = .52, SD = .13), and high-stake (M = .52, SD = .13) trials ($t(84) = -.25$, $d = -.03$, $p = .803$, 95% CIs [-.03, .03]) for the children. This suggests that children did not show a stakes effect like adults (Figure 12a).

When I 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 [-.18, -.01]), and high-stake trials ($t(30.00) = -4.35$, $d = 1.21$, $p < .001$, 95% CIs [-.27, -.10]).

I next tested whether an effect of stake on model-based decision-making might emerge with age for the children. Therefore, I correlated the model-based contribution parameters for the children's low and high-stake trials separately with age and controlled the age-related slopes during high and low-stake trials for the correlation between the two contribution parameters (Figure 12b). The difference between the slopes was not significant (z = -0.50, p = .616). Thus, a stakes effect was not apparent in the children's behavior, suggesting that this ability may emerge later during development (Figure 12c).

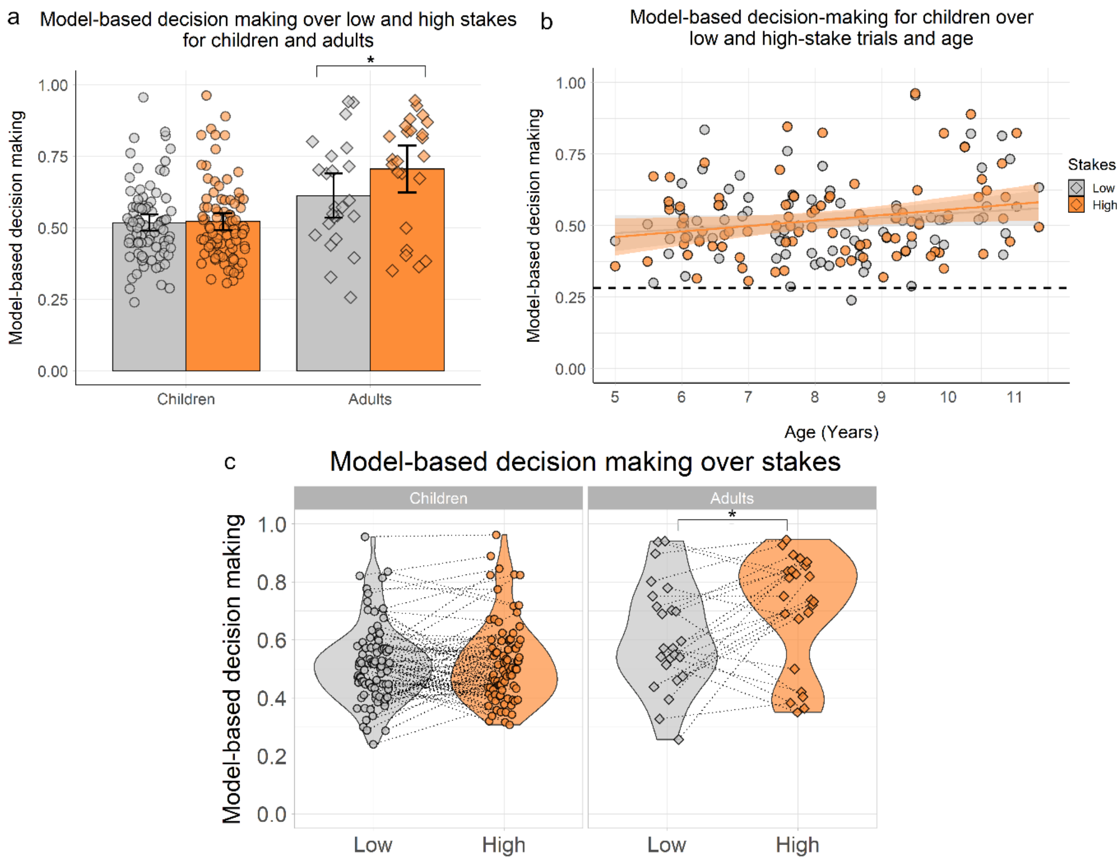


Figure 12. Model-based decision-making over stakes for adults and children.

a) Adults displayed a significantly higher degree of model-based decision-making for the high-stake trials. b) While children did not show a difference in the degree of model-based decision-making over stakes, this did not change over age. 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 < .05; **p < .01; ***p < .001.

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

To complement the computational modeling analyses, I 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. I used the same regression method as in an earlier task version (Kool et al., 2016). Using this method, the model-based component consists of a main effect of the previous reward on the probability 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). The previous reward refers to the continuous points won by the participant on the previous trial. Starting state similarity refers to whether the current starting state (the rocket pair) is the same as in the previous trial. The influence of previous reward on staying behavior approximates the transfer of experience from one starting state to the other. On the other hand, 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 (Figure 13a).

First, I 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 of stay, indicating a model-based component (beta =

.12, se = .02, z = 5.56, p < .001). The interaction between previous reward and starting state similarity was not significant, showing that previous reward increased the probability of staying for both starting states similarly (beta = -.003, se = .02, z = -.14, p = .892). In addition, there was a main effect of starting state (beta = .05, se = .02, z = 2.35, p = .02). Thus, these results suggest that children could generalize successfully over starting states and indicated a model-based component in their behavior (Figure 13b).

For adults, there was also a main effect of reward on staying probability (beta = 1.09, se = .05, z = 22.81, p < .001). There was no main effect of starting state (beta = .06, se = .05, z = 1.44, p = .149), however, there was a small but significant interaction between starting state and previous reward (beta = .10, se = .05, z = 2.22, p = .026) (Figure 13c). I also included a group term in the models to compare children and adults. the model-based predictor, previous reward, remains significant for the whole sample (beta = 0.12, se = 0.02, z = 5.55, p < .001). I 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 < .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 < .001). Adults also had a higher raw behavioral stay probability overall than the children (F(1,12631) = 120.9, p < .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.

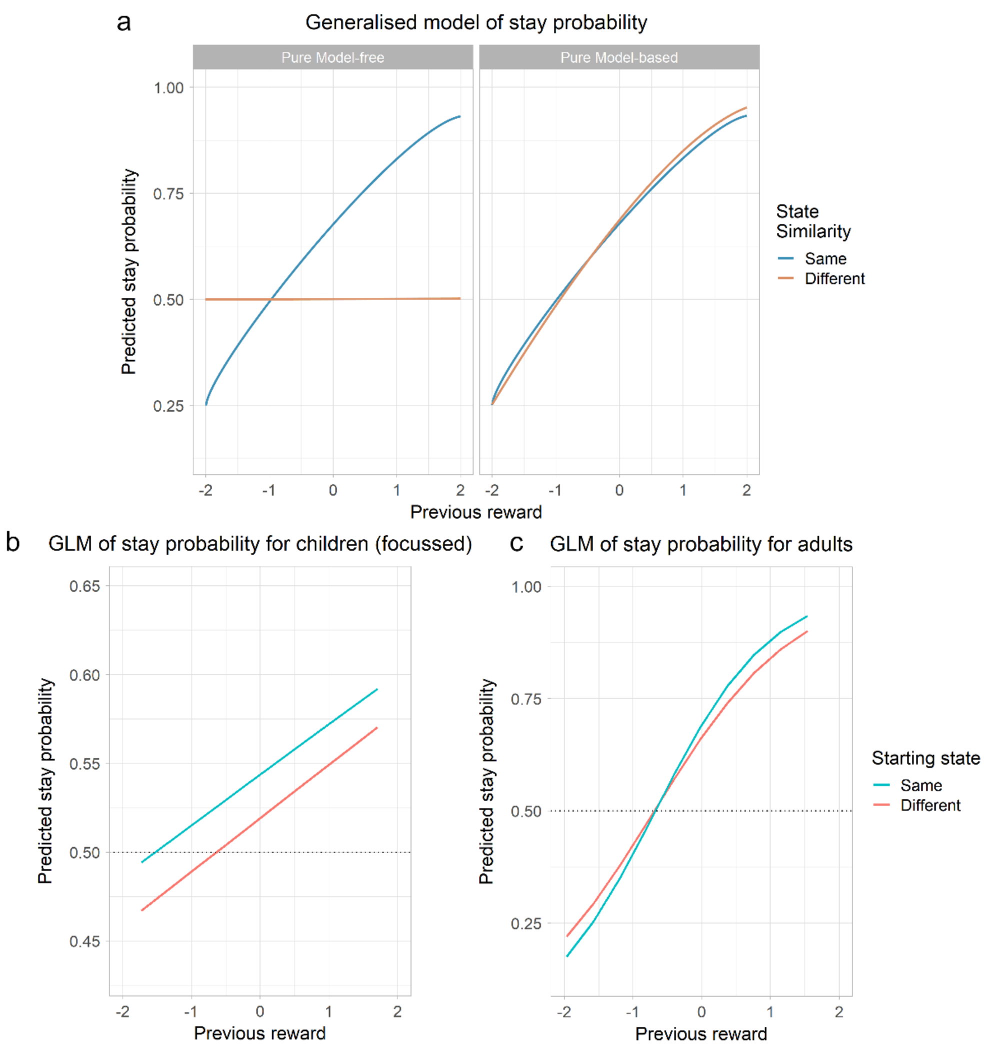


Figure 13. Model-free and model-based contributions to stay probability.

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 of staying 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).

2.4.5 Best-fitting behavioral models for children and adults

Next, I conducted a nested model selection to find the best model to separately predict stay probability for children and adults. In a previous logistic regression model, additional predictors were included alongside previous reward (the model-based component) and starting state similarity (same or different spaceship pairs) to approximate the computational models more closely. Namely, the difference in available reward 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, I also included age for the children. For both children and adults, I included a null model with only an intercept 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; Table 4). Previous reward had a significant main effect on staying probability (beta = .12, se = .02, z = 5.60, $p < .001$), while age was not a significant main effect (beta = -.00, se = .04, z = -.04, p = .967), but the interaction between previous reward and age was significant (beta = .070, se = .02, z = 3.17, p = .002) (Figure 14a). Thus, previous reward had a main effect on staying probability, indicating a significant model-based effect on the children's choice behavior. The positive interaction shows that the influence of previous reward on staying probability increases with age.

Table 4. Generalized linear model for children.

	Dependent variable:
	Stay probability (planet) Linear Mixed Effects
Previous reward	0.119***
	(0.077,0.160)
Age	-0.002
	(-0.084,0.081)
Previous reward x Age	0.067***
	(0.026,0.109)
Constant	0.122***
	(0.039,0.205)
Number of Participants	85
sd(Participant)	0.337
N	9456
Note:	*p<0.1; **p<0.05; ***p<0.01

For adults, the best-fitting model included previous reward, starting state, and stake, as well as their interactions (AIC weight (model probability) = 0.83, Table 5). There were significant fixed effects of previous reward (the model-based component) (beta = 1.14, se = .05, z = 22.78, p < .001) and stake (beta = 0.22, se = .05, z = 4.88, p < .001). Additionally, the interaction between previous points and stake were significant, indicating a stakes effect (beta = .35, se = .05, z = 7.08, p < .001), (Figure 14b). The interactions between previous points and state similarity was also significant (beta = .13, se = .05, z = 2.56, p = .010), and the three-way interaction between previous points, starting state and stake (beta = .11, se = .05, z = 2.25, p = .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.

Table 5. Generalized linear model for adults.

	Dependent variable:
	Stay probability (planet) Linear Mixed Effects
Previous reward	1.139***
	(1.041,1.237)
Starting State	0.080*
	(-0.008,0.169)
Stake	0.221***
	(0.132,0.310)
Previous reward x Same	0.125**
	(0.029,0.221)
Previous reward x Stake	0.347***
	(0.251,0.443)
Starting state x Stake	0.001
	(-0.088,0.090)
Previous reward x Starting State x Stake	0.110**
	(0.014,0.206)
Constant	0.778***
	(0.554,1.001)
Number of Participants	24
sd(Participant)	0.512
N	3177
Note:	*p<0.1; **p<0.05; ***p<0.01

To compare children and adults directly, I ran models that included the whole sample of participants and looked at group-based differences.

First, I assessed group-based differences in the model-based predictor. When I include group into the model (stay ~ previous reward * starting state similarity * Group + (1|ID)), the model-based predictor, previous reward, remains significant for the whole sample (beta = 0.12, se = 0.02, z = 5.55, $p < .001$). I also see a significant main effect of group on the probability to stay (beta = 0.44, se = 0.10, z = 4.41, $p < .001$), with the adults scoring higher overall. In addition, there is an interaction between previous reward and group, (beta = 0.98, se = 0.5, z = 18.67, $p < .001$), showing that adults show a stronger effect of the model-based predictor on stay probability. There is also a main effect of starting state on the probability to stay overall, (beta = 0.05, se = 0.02, z = 2.34, p = 0.019), and an interaction for previous reward and starting state for the adults (beta = 0.11, se = 0.05, z = 2.07, p = 0.039). Based on this analysis, the significant interaction between group and previous reward indicates that the model-based predictor had a stronger effect for the adults.

When I further included stake in the model, I 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, $p < .001$), indicating that adults showed more model-based control during high stake trials. There was also a significant interaction of stake and group (beta = 0.17, se = 0.05, z = 3.47, p = 0.001), and the interaction between previous points and groups remained significant (beta = 1.03, se = 0.05, z = 18.89, $p < .001$), as well as the main effect of previous points (beta = 0.12, se = 0.02, z = 5.54, $p < .001$), and the main effect of group (beta = 0.47, se = 0.10, z = 4.64, $p < .001$), indicating that adults showed higher stay probability overall and higher stay probability for the model-based predictor, but that the model-based predictor was still significant for children and adults alike. Additionally, there was a main effect of starting state (beta = 0.05, se = 0.02, z = 2.32, p = 0.020), a three-way interaction between previous reward, starting state and group (beta = 0.13, se = 0.05, z = 2.39, p = .017), and a four-way interaction between previous reward, starting state, stake and group (beta = 0.11, 0.05, z = 2.16, p = .031), indicating that adults were more likely to stay if the starting state was the same, especially for high-stake trials and after larger previous rewards. Thus, I see a stake effect repeated for the adults using the regression methods and an absence of a stakes effect for the children.

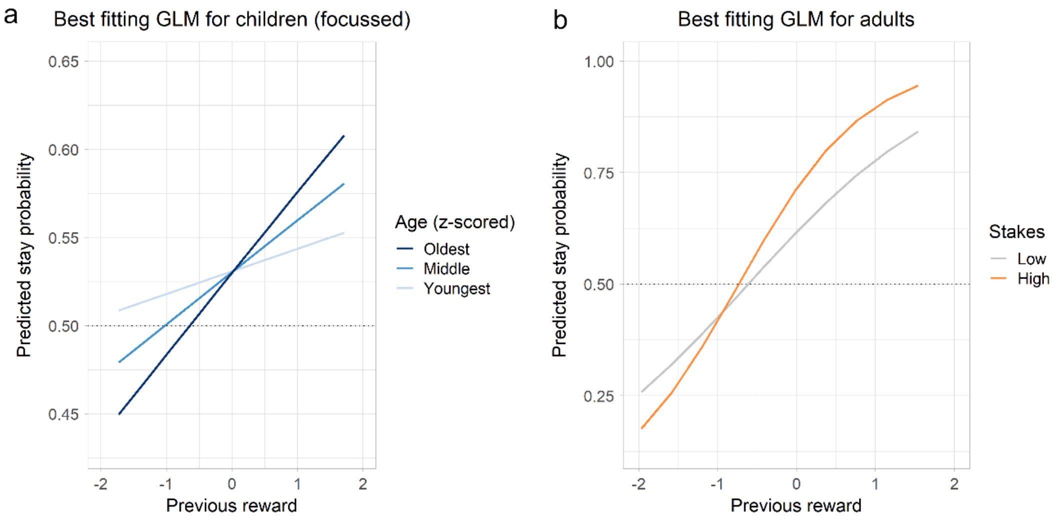


Figure 14. Best-fitting generalized linear mixed models of stay probability for children and adults.

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 and 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%).

2.5 Discussion

I investigated the development of model-based decision-making and how this is used adaptively across contexts in children aged 5-11. I report that when using a two-step task that encourages computationally costly decision-making strategies, children aged 5-11 years demonstrated model-based decision-making. This finding was supported by both computational and behavioral measures of model-based decision-making. Crucially, I found that even five-year-old children showed robust model-based decision-making, while the degree to 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 as young as five can engage in sophisticated decision-making strategies on a sequential choice task but that the optimal arbitration between strategies undergoes further development.

The finding that children younger than 12 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 studies revealed a developmental increase in model-based decision-making from childhood to adulthood. However, they also indicated that children 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 five years old. I attribute the discrepant results 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 the Discussion, section 5.2 On the current and previous contrasting findings of model-based decision-making in childhood).

Thus, the high complexity and uncertainty in tasks 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 increased 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, the current 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 can engage in sophisticated decision-making strategies when the task allows for this.

Contrarily, I 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 specific situation might be the actual 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, even up to late adolescence, children might be less able to detect and assign values to relevant environmental cues than adults, leading them to respond similarly to rewards of different magnitudes (Davidow et al., 2018; Insel et al., 2019). However, while the absence of metacontrol may reflect a genuine developmental effect, 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., 2022; Veselic et al., 2021). Future work may wish to use incentives that are even more salient to the present age group 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 EFs or memory storage or manipulation, remains unclear.

While model-based decision-making was present throughout the age ranges in this 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., 2015; Otto, Raio, et al., 2013; 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 essential 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. However, the behavioral data indicated a higher probability of staying with higher previous reward, indicating 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, the finding that children were less likely to repeat a visit indicates one of the most considerable behavioral differences between children and adults. This might be due to children being less successful in exploiting 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. Thus, while children showed robust markers of model-based decision-making in that their behavior did not differ across starting states, their behavior differed from adults, mainly due to being less likely to repeat visits to the same planet.

Additionally, I 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 to limit timed-out trials for younger developmental samples.

Lastly, while 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 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.

In summary, this study demonstrates the presence of sophisticated value-based decision-making strategies during childhood. I 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 could 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 metacontrol, 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.

2.6 Supplementary Materials

2.6.1 Dual-reinforcement learning model

To estimate the degree of model-based decision-making participants employed in this paradigm, I fitted an established dual-system reinforcement learning model (Daw et al., 2011; Gläscher et al., 2010; Kool et al., 2016, 2017) to their behavior. The paradigms consist of four states across two stages (the two pairs of spaceships and the two planets), with two available actions at the first-stage states between the spaceships (a_A and a_B) and one action at the second-stage state to collect the treasure (a_C). The reinforcement-learning model consists of a model-based and a model-free system that both learn different values for actions and states, denoted as $Q(s, a)$, which map each state-action pair to its expected discounted future return. On trial t, the first-stage state is denoted by $s_{1,t}$, the second-stage state by $s_{2,t}$, the first and second stage actions by $a_{1,t}$ and $a_{2,t}$, and the first and second stage rewards as $r_{1,t}$ (which is always zero, since only on the second stage reward is attained) and $r_{2,t}$.

Model-free agent. The model-free agent relies on the state-action-reward-state-action (SARSA) temporal difference learning algorithm, which uses reward prediction errors, the learning rate, and the eligibility trace to update the values for each state-action pair *(s, a)* at stage *i* and trial *t* according to:

$$Q_{MF}(s,a) = Q_{MF}(s,a) + \alpha\delta_{i,t}e_{i,t}(s,a)$$

where

$$\delta_{i,t} = r_{i,t} + Q_{MF}(s_{i+1,t}, a_{i+1,t}) - Q_{MF}(s_{i,t}, a_{i,t})$$

Is the reward prediction error for trial *t* at stage *i*, *a* is the learning rate parameter, which determines to which degree new information is incorporated, and $e_{i,t}(s,a)$ is an eligibility trace parameter, and which is set equal to 0 at the beginning of each trial and updated according to:

$$e_{i,t}(s_{i,t}, a_{i,t}) = e_{i-1,t}(s_{i,t}, a_{i,t}) + 1$$

before the *Q* value update. The eligibilities of all state-action pairs are then decayed by *λ* after the update.

For the current paradigm, this learning rule applies in the following way. The reward prediction error is different for the first two levels of the paradigm. Since at the first stage where they choose the spaceships, there is no reward, $r_{1,t}$ is always zero. The reward prediction at the first stage is instead driven by the value of the selected second stage action $Q_{MF}(s_{2,t},a_{2,t})$:

$$\delta_{1,t} = Q_{MF}(s_{2,t}, a_{2,t}) - Q_{MF}(s_{1,t}, a_{1,t})$$

This means that the predicted reward from choosing the spaceships is tied to the reward attained at the planet stage. Since there is no third stage, the second stage prediction error is driven by the reward $r_{2,t}$:

$$\delta_{2,t} = r_{2,t} - Q_{MF}(s_{2,t}, a_{2,t})$$

Both the first- and second-stage values are updated at the second stage, with the first-stage values receiving a prediction error that is down-weighted by the eligibility trace decay *lambda*. When *lambda* = 0, only the values of the current state get updated, rather than the values in the past.

Model-based agent. The model-based agent uses the same reward prediction errors and learning rate as the model-free agent, but in addition, uses the transition map of the paradigm to calculate values of each choice. For this paradigm, it means that a model-based agent, but not a model-free agent, can generalize over choices in the two different starting states. To get an intuition for how this leads to different forms of behavior, say, for example, that a participant chooses the blue spaceship, which then transitions to the red planet, and this leads to a large reward. In the next trial, the participant is presented with the other starting state, the one that does not have the previously chosen blue spaceship. Now, the model-based system will realize that the orange spaceship also transitions to the red planet, and because it has just learned that this planet has become better, it will increase its preference for this choice option. A model-free agent is not able to make such generalizations since it relies on direct learning from action-reward contingencies. Therefore, it will not be more likely to pick the orange spaceship over the light blue spaceship in the other starting state. In short, a model-free agent would generate four separate values for all the spaceships, while a model-based agent would only generate two, correctly learning that two spaceships transition to the same planet.

The model-based values are defined in terms of the Bellman's equation, which specifies the expected values of each first-stage action using the transition structure *P*, which means knowing how the spaceships transition to the planets, and which is assumed to be known to the agent:

$$Q_{MB}(s_A, a_j) = P(s_B|s_A, a_j) \max_{a \in \{a_A, a_B\}} Q_{MF}(s_B, a) + P(s_C|s_A, a_j) \max_{a \in \{a_A, a_B\}} Q_{MF}(s_C, a)$$

where I have assumed these are recomputed at each trial from the current estimates of the transition probabilities and second-stage reward values.

Decision rule. To connect the model-based and model-free values to choices, the Q-values are then mixed according to a weighting parameter *w*:

$$Q_{net}(s_A, a_j) = wQ_{MB}(s_A, a_j) + (1 - w)Q_{MF}(s_A, a_j).$$

Where a value closer to 1 means the agent is more model-based, and a value closer to 0 means the agent is more model-free. To accommodate the stake manipulation, I defined two different weights that operated on different trial types. I set $w = w_{low}$ on low stake trials and $w = w_{high}$ on high stake trials.

In the second stage, the decision is made using only the model-free values. I used the Softmax rule to translate the weighted Q-values to actions. This rule computes the probability for an action, reflecting the combination of the model-based and model-free action values weighted by an inverse temperature parameter. At both states, the probability of choosing action *a* on trial *t* is computed as:

$$P(a_{i,t} = a|s_{i,t}) = \frac{exp(\beta[Q_{net}(s_{i,t}, a) + \pi \cdot rep(a) + \rho \cdot resp(a)])}{\sum_{a'} exp(\beta[Q_{net}(s_{i,t}, a') + \pi \cdot rep(a') + \rho \cdot resp(a\prime)])}$$

where the inverse temperature *b* determines the randomness of choice or the exploitation/exploration trade-off. Specifically, when *b* approaches infinity, the probability of choosing the action with the highest expected value tends to be 1, whereas, for *b* approaching 0, the probabilities over actions become equally likely across all options. The indicator variable *rep(a)* is defined as 1 if *a* is a first-stage action (choosing a spaceship) and is the same one as was chosen in the previous trial, so the participant chose the same rocket, zero otherwise. Multiplied with the 'stickiness' parameter *p*. This captures the degree to which participants show perseveration (when $p > 0$) or switching ($p < 0$) at the first stage. The indicator variable *resp(a)* is defined as 1 if *a* is a first-stage action selecting the same response key as the key that was pressed on the previous trial, zero otherwise. Multiplied with the response stickiness parameter *r*, this captures the degree to which participants repeated ($r > 0$) or alternated ($r < 0$) key presses at the first stage (e.g., whether they pressed the left key twice in a row). These two stickiness parameters were used since the locations of the spaceships changed per trial, and participants could therefore show perseveration or alternation bias towards the spaceships, button presses, or both.

2.6.2 Model-fitting procedure

I used maximum *a posteriori* estimation, implemented using the *mfit* toolbox (Gershman, 2018), to fit the parameters for the 6 (dual-systems reinforcement learning model with one mixing weight) and 7-parameter (dual-systems reinforcement learning model with two mixing weights per stake) computational models to observed data. To avoid local optima in the estimation solution, the optimization was run 100 times for each participant with randomly selected initializations for each parameter.

2.6.3 Parameter recovery

To test whether the 7-parameter reinforcement learning model could reliably identify the contributions of both model-free and model-based decision-making on the task, I conducted parameter recovery for the 7-parameter model by running the generative version of the model for 500 agents and for 100, 140, and 200 trials. For each agent, I randomly sampled the initial parameters from uniform distributions: for all parameters bounded between 0 and 1 (learning rate a, eligibility trace l, *w*-low, *w*-high), I used U(0,1), for inverse temperature b U(0,2), and for the stickiness parameters π and r I used U(-0.5,0.5) (Bolenz et al., 2019; Kool et al., 2016). Next, I used the same model-fitting procedures as for the participant data to estimate the model parameters of the simulated data.

For 100 trials, I found substantial correlations between the estimated parameters for w-low (r = .61) and w-high (r = .60). For 140 trials, the correlations were slightly stronger (w-low: r = .62, w-high: r = .66), similar to the estimated parameters for 200 simulated trials (w-low: r = .69, w-high: r = .75). This indicates that for the trial ranges present, I could extract meaningful parameter estimates for the model-based parameters across stakes.

For the other parameters, for 100 trials I found: β: r = .87, α: r = .79, λ: r = .45, π: = .44, ρ: = .58. For 140 trials: β: r = .90, α: r = .83, λ: r = .53, π: = .52, ρ: = .67, and for 200 trials: β: r = .92, α: r = .87, λ: r = .54, π: = .54, ρ: = .71.

2.6.4 K-fold cross-validation

I used k-fold cross-validation to test how accurate the models predicted behavior for children and adults. I conducted the k-fold cross-validation for each model separately and assessed model performance in four ways for children and adults. First, I evaluated the mean accuracy of each reinforcement learning model and tested these against the accuracy of the random model. Second, I conducted Bayesian model selection for the four models based on the predictive accuracies as established with the k-fold cross-validation. Third, I compared model AICs. Lastly, I used an additional parameter recovery analysis for the winning model to see if I could recreate human behavior and recover the same parameter solutions for the participants.

The procedure for the k-fold cross-validation was as follows. For the full task, trials consisted of four blocks of 35 trials each. For every participant, I created four different combinations of training blocks (3) and left-out blocks (1). I fitted the model to the three training blocks and then used the parameter solutions to predict decisions for the left-out block. I evaluated the likelihood of the choice given the model (parameters) to assess the predictive accuracy. This provided the model accuracy measure for the four models, 1) the random decision model, 2) the simplified reinforcement learning model (henceforth 3-parameter model), 3) the 6-parameter stake-agnostic dual-systems model (henceforth 6-parameter model), and 4) the 7-parameter metacontrol dual-systems model (henceforth 7-parameter model). To test whether the model predicted choice behavior significantly above chance level, I compared the model accuracies of the three reinforcement learning models to the accuracy of the random model.

For the children, both the 6-parameter model (*M* = .5347, *SD* = .0868) and the 7-parameter model (*M* = .5345, *SD* = .0864) explained behavior significantly better than the random model (M = 0.5000, SD = 0), (6-parameter model: *t*(84) = 3.68, *d* = 0.40, *p* < .001, 95% CIs [.516, .553]; 7-parameter model: *t*(84) = 3.68, *d* = 0.40, *p* < .001, 95% CIs [.516, .553]). The 3-parameter model (M = 0.4939, SD = 0.0209), predicted behavior significantly worse than the random model (t(84) = -2.67, d = 0.29, 95% CIs [0.4894, 0.4985]). There was no significant difference in model accuracy between the 6-parameter and the 7-parameter models (*t*(84) = 0.50, *d* = 0.05, *p* = .656, 95% CIs [-.0005, .0008]), but the 6-parameter model did explain behavior significantly better than the 3-parameter model (t(84) = -4.58, d = 0.50, 95% CIs [-0.058, -0.023]), and so did the 7-parameter model (t(84) = -4.58, d = 0.50, 95% CIs [-0.058, -0.023]).

For the adults, I found that the 6-parameter (*M* = .5442, *SD* = .0589), the 7-parameter model (*M* = .5413, *SD* = .0572), and the 3-parameter model (M = 0.5314, SD = 0.0555) explained behavior significantly above chance level (6-parameter model: *t*(23) = 3.67, *d* = 0.75, *p* = .001, 95% CIs [.519, .569]; 7-parameter model: *t*(23) 3.54, *d* = 0.72, *p* = .002, 95% CIs [.517, .565]; 3-parameter model: t(23) = 2.78, d = 0.57, p = .011, 95% CIs [.5080, .5549]). The 3-parameter model explained behavior significantly worse than both the 6-parameter (t(23) = -2.61, d = 0.53, 95% CIs [-0.023, -0.003]), and the 7-parameter model (t(23) = -2.11, d = 0.43, 95% CIs [-0.020, -0.0002]). There was a significant difference in model accuracy between the 6-parameter and the 7-parameter model for the adults (*t*(23) = 2.85, *d* = 0.58, *p* = .009, 95% CIs [0.001, .005]). Overall, the 6-parameter model numerically had the highest accuracy, but there was no significant difference between the 6-parameter and the 7-parameter models for adults.

2.6.5 Bayesian model comparison

Next, I conducted Bayesian Model Comparison (BMS) using the bms function in the Matlab mfit package (Gershman, 2016, 2018; Stephan et al., 2009). I used the predictive accuracy of each model as established via the k-fold cross-validation and reported the winning model based on the exceedance probabilities for the children and adults separately.

For the children, the 6-parameter model had the highest exceedance probability (EP = 0.307), while the 7-parameter model came second (EP = 0.280), then the 3-parameter model (EP = 0.249), and lastly the random model (EP = 0.164).

For the adults, the 7-parameter model had the highest exceedance probability (EP = 0.309), next the 6-parameter model (EP = 0.297), then the 3-parameter model (EP = 0.226), and lastly the random model (EP = 0.169).

Next, I compared model AICs for both the children and adults separately. First, I assessed the model with the minimum AIC and the delta AIC values with the other models. For the children, the model with the lowest AIC was the 6-parameter model (AIC = 163.04). The 7-parameter model had the second lowest (AIC = 164.86), and the delta AIC between them was negligible (ΔAIC = 1.82). The 3-parameter model had the next lowest (AIC = 172.69), and the delta AIC with the 6-parameter model was substantial (ΔAIC = 9.65). Lastly, the random model (AIC = 173.44) had the most considerable AIC difference from the 6-parameter model (ΔAIC = 10.40).

For the adults, the model with the smallest AIC was the 3-parameter model (AIC = 149.77). The 6-parameter model had the second-lowest AIC (AIC = 151.92). The difference between the 3-parameter and 6-parameter models was small (ΔAIC = 2.14). The 7-parameter model was the third model with the lowest AIC (AIC = 153.67). The difference between the 7-parameter and 3-parameter models was small (ΔAIC = 3.89). Lastly, the random model (AIC = 190.71) had the most considerable AIC difference with the winning 3-parameter model (ΔAIC = 40.94). The difference between the 6-parameter model and the random model was also substantial (ΔAIC = 38.80).

Overall, for children and adults, the 6-parameter stake-agnostic dual-systems model comes out as the best-fitting model for the data. I, therefore, use this model as the winning model. The 7-parameter metacontrol dual-systems model also has a good fit, performing better than the 3-parameter simplified reinforcement learning model in the mean accuracy of model prediction and the Bayesian model comparison. I, therefore, use both these models in the results going forward.

2.6.6 Qualitative model validity

To test whether the winning model successfully captured human behavior, I also conducted an additional parameter recovery analysis using the participant parameter solutions rather than simulated data. This allows us to test whether the models can successfully capture human behavior and whether unique parameter solutions are recoverable. The current approach to this was as follows; I used the parameter solutions from the participants, children, and adults ran separately to simulate new behavior using the 6-parameter stake-agnostic dual-systems model. Next, I fitted the model to the simulated data and extracted the parameter solutions again. I next correlated the initial and final parameter solutions to each other. A strong correlation would indicate that the model can both recover the unique parameter solutions of participants reliably and able to simulate human behavior.

For the children, all parameter solutions had significant positive correlations with the initial and recovered parameter solutions (inverse temperature: $r = 0.74$; learning rate, $r = 0.77$; eligibility trace: $r = 0.27$; model-based decision-making weighting parameter (w): $r = 0.55$; rocket stickiness parameter: $r = 0.78$; key stickiness parameter: $r = 0.78$). For the adults, all initial and recovered parameters were also significant (inverse temperature: $r = 0.90$; learning rate: $r = 0.82$; eligibility trace: $r = 0.26$; model-based decision-making weighting parameter (w): $r = 0.66$; rocket stickiness parameter: $r = 0.70$; key stickiness parameter: $r = 0.75$). Thus, the winning model could simulate human behavior and recover the parameter solutions.

Lastly, I included a qualitative model assessment of the winning model by comparing the human behavior to simulated behavior via the number of points won during the task. I (1) used the parameters obtained from the human participants to simulate new behavior across ten iterations, and (2) compared the behavioral performance of the human participants to the simulated participants. Using an ANOVA, I tested whether there was a difference in the behavior of the humans and simulated participants across all iterations (where one iteration was data from the human participants, Figure 15), and I found no difference between them ($F(9,1080) = 0.77$, $p = .647$, $\eta^2 = 0.006$). I also tested this separately for each iteration (we compared the iteration with human participants to each simulated iteration separately). Here I find that for nine of the iterations there was no significant difference between the humans and simulated behavioral performance (it2: $F(1,216) = 3.24$, $p = .073$, $\eta^2 = 0.015$; it4: $F(1,216) = 3.13$, $p = .078$, $\eta^2 = 0.014$; it5: $F(1,216) = 2.54$, $p = .112$, $\eta^2 = 0.012$; it6: $F(1,216) = 3.41$, $p = .066$, $\eta^2 = 0.015$; it7: $F(1,216) = 3.24$, $p = .073$, $\eta^2 = 0.015$; it8: $F(1,216) = 2.73$, $p = .100$, $\eta^2 = 0.012$; it9: $F(1,216) = 2.82$, $p = .094$, $\eta^2 = 0.013$; it10: $F(1,216) = 3.35$, $p = .069$, $\eta^2 = 0.015$). For one iteration, the difference was significant (it3: $F(1,216) = 4.34$, $p = .038$, $\eta^2 = 0.020$). Combined with the previous results, I conclude that the model seems capable of reproducing human behavior.

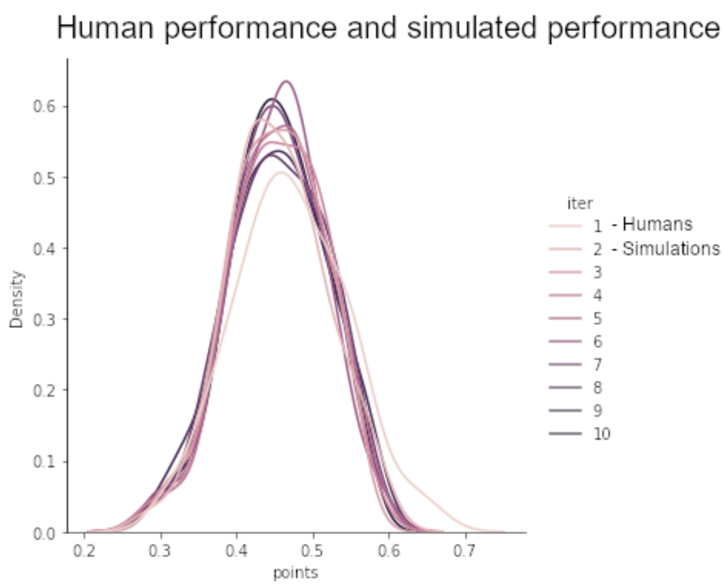


Figure 15. Qualitative model comparison.

Iteration one (the lightest color) indicates human performance on the task. The other iterations (2-10) show the performance of the simulated agents based on human parameters.

2.6.7 Trial-by-trial analyses of reaction time and performance

2.6.7.1 Reaction Time over Trials

I conducted regression analyses to investigate whether reaction time changed over the course of the task, using reaction time as the outcome variable and trial as the predictor with a random intercept per participant for children and adults separately (RT ~ Trial + (1|Participant)). For children there was no change in reaction time over trials, ($b < .0001$, $se = 0.0001$, $p = .969$, 95% CIs [-.02, .02], (Figure 16 top, first plot)), while adults became significantly faster over trials, ($b = -.001$, $se = 0.0001$, $p < .001$, 95% CIs [-.12, -.06], (Figure 16 bottom, first plot)).

I next ran a model including stakes (RT ~ Trial * Stake + (1|Participant)). For children, there was no effect of stake or a trial by stake interaction, (trial: $b = .0002$, $se = .0002$, $p = .301$, 95% CIs [-.01, .04], stake: $b = .040$, $se = .022$, $p = .064$, 95% CIs [-.01, .05], stake x trial: $b = .0004$, $se = .0003$, $p = .135$, 95% CIs [-.06, .01]). For the adults, when including stake, there was also no significant effect of stake, or a stake by trial interaction, but the effect of trials remained (trial: $b = .0006$, $se = .0002$, $p = .001$, 95% CIs [-.11, -.03], stake: $b = .0182$, $se = .021$, $p = .383$, 95% CIs [-.07, .04], stake x trial: $b = .0003$, $se = .0003$, $p = .187$, 95% CIs [-.10, .02]).

Lastly, I investigated whether there was a stake effect on reaction time, for example, if participants slowed down on high-stake trials. I conducted a regression analysis for reaction time, using stake and trial predictors. There was no significant main effect of stake for either children, ($b = .04$, $se = 0.02$, $p = .064$), or adults, ($b = .02$, $se = .02$, $p = .383$).

Important to note is that the participants also had time to make their decision when observing the treasure amplifier, which was always shown for 2 seconds. Thus, participants could have already made their decision before seeing the spaceship pairs.

2.6.7.2 Performance over trials

To assess potential training or fatigue effects over trials, I analyzed the performance in the form of points won over trials for children and adults. I also plot performance over trials separately for children and adults, and I have split the two samples into three age quantiles to assess if there were age-related changes. I have plotted all trials, which includes the timed-out trials (where rewards won = 0), to indicate broader variability in the data (see Figure 16 top, the second plot for the children, and Figure 16 bottom, the second plot for the adults).

I ran a linear model investigating whether trial number predicted rewards won, indicating that the participants became better at the task over time. There was a positive association between trial, and points won for the children, showing that with higher trial numbers, children won more rewards (beta = 0.004, se = 0.001, t = 6.48, $p < .001$). For the adults, this was also positively significant, (beta = 0.003, se 0.002, t = 2.19, p = 0.029). When I ran a model including children and adults to see if there was potentially a group interaction, both trial and group were significant predictors, but there was no significant interaction between them (beta = -0.0008, se 0.2, t = -0.49, p = .623). Increasing trial numbers still predicted more rewards won (beta = 0.004, se = 0.001, t = 6.53, $p < .001$), and adults won more rewards overall (beta = 0.9, se = 0.1, t =

5.68, p < .001). Thus, both groups seemed to perform better over time, with children improving more than adults but adults performing better overall. This suggests that there are training effects rather than fatigue effects. Do note that participants were allowed three breaks throughout the main section of the task for as long as they wanted after completing each block of 35 trials.

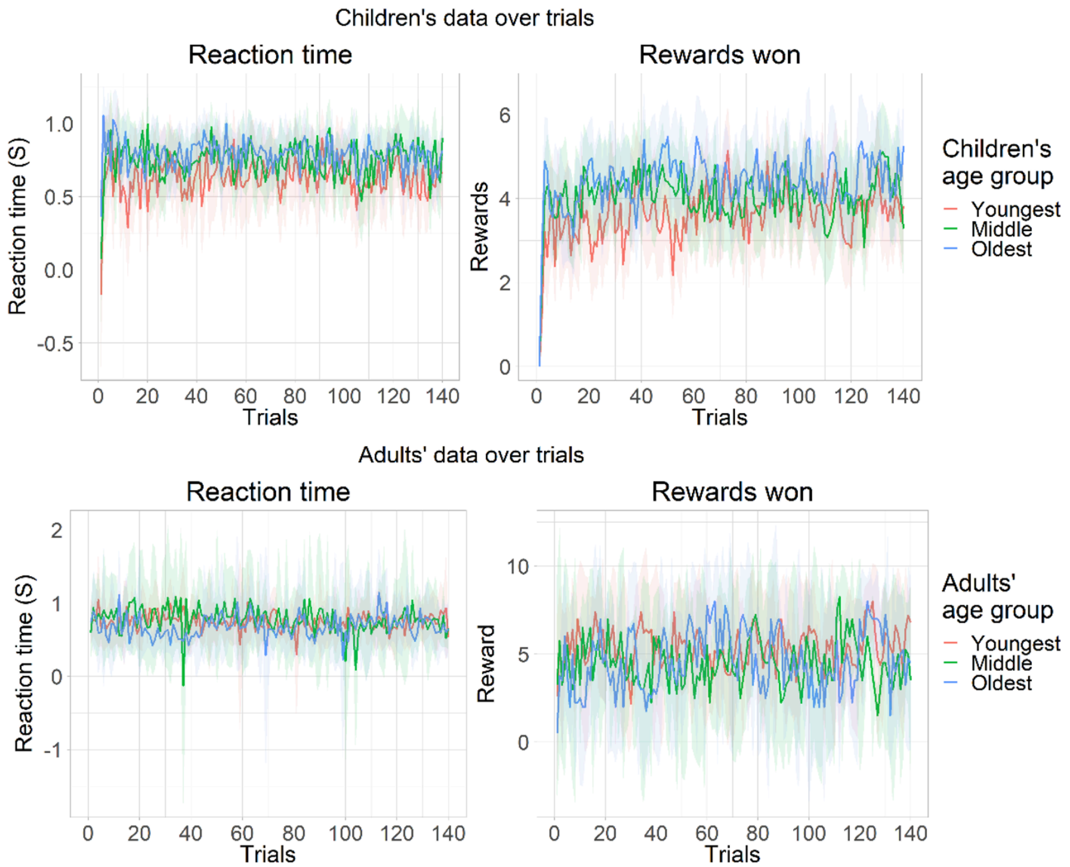


Figure 16. Reaction time and performance over trials.
Performance in the form of points won over trials for children (top) and adults (bottom). Both groups have been split into three age groups of equal numbers of participants for visualization. Error bars depict 95% confidence intervals, and shaded areas around regression lines indicate the standard error of the mean. Shaded areas represent the 95% confidence intervals of the means.

2.6.8 Potential model-based effects on reaction time

I ran two regression models to investigate whether reaction times could serve model-based decision-making. First, I ran models where I tried to predict reaction times on the current trial based on whether the previous trial had been a "win" (same definition as above), whether the current starting state was the same or different and whether they chose to "stay" or "switch" (e.g., lmer(RT ~ same * win * stay + (1 | ID)). This model would tell us whether reaction times might significantly slow down after large reward prediction errors and if this differed by state similarity.

Next, I ran models where I tried to predict behavioral staying ("stay" or "switch") based on reaction time on the current trial, whether the previous trial had been a "win" and whether the current starting state was the same or different, (e.g., glmer(stay ~ RT * same * win + (1 | ID)). This model would tell us whether reaction times predicted behavioral staying as a function of starting state and previous win. I ran both these models for the children and adults separately.

For the children, for the first model, there was no significant effect of either starting state (beta = 0.002, se = 0.008, t = 0.23, p = 0.818), behavioral staying (beta = -0.006, se = 0.011, t = -0.53, p = 0.595), or previous win (beta = -0.011, se = 0.011, t = -0.97, p = 0.330), on reaction time. Thus, reaction time during the first stage for children did not seem to be significantly affected by the predictors, see Figure S5a, the first plot.

For the second model, there was a significant interaction between reaction time and previous win (beta = 0.36, se = 0.10, z = 3.59, $p < .001$), showing that higher reaction times after a win were associated with a higher stay probability, but that this was not affected by starting state (same: beta = 0.01, se = 0.07, z = 0.19, p = 0.847). Thus, after a win, longer reaction times were associated with a higher probability of staying for the children, see Figure S5a, the second plot.

For the adults, for the first model, there were significant negative main effects of starting state (beta = -0.03, se = 0.01, t = -2.77, p = .006), and a significant negative three-way interaction between starting state, behavioral staying, and a previous win, (beta = -0.06, se = 0.02, t = -2.85, p = 0.003), showing that adults responded faster on trials in the same starting state as the previous one, especially if they decided to stay after a previous win. There was also a positive interaction between starting state similarity and previous win (beta = 0.04, se = 0.02, t = 2.32, p = 0.020), showing that adults responded slower when they were in the same starting state after a win, which could reflect their decision time on whether to stay or switch.

For the second model, there was a main significant effect of previous win, (beta = 2.14, se = 0.22, z = 9.51, p < .001), and a significant interaction between starting state and previous win (beta = 0.44, se = 0.22, z = 1.99, p = .047). This suggests that adults were more likely to stay after a previous win and that they were more likely to stay after a win in the same starting state. Thus, there were no significant effects of reaction time on staying probability for the adults.

Thus, there were some indications that reaction times reflected processing time after previous wins and depending on starting state, but these were not the same between children and adults.

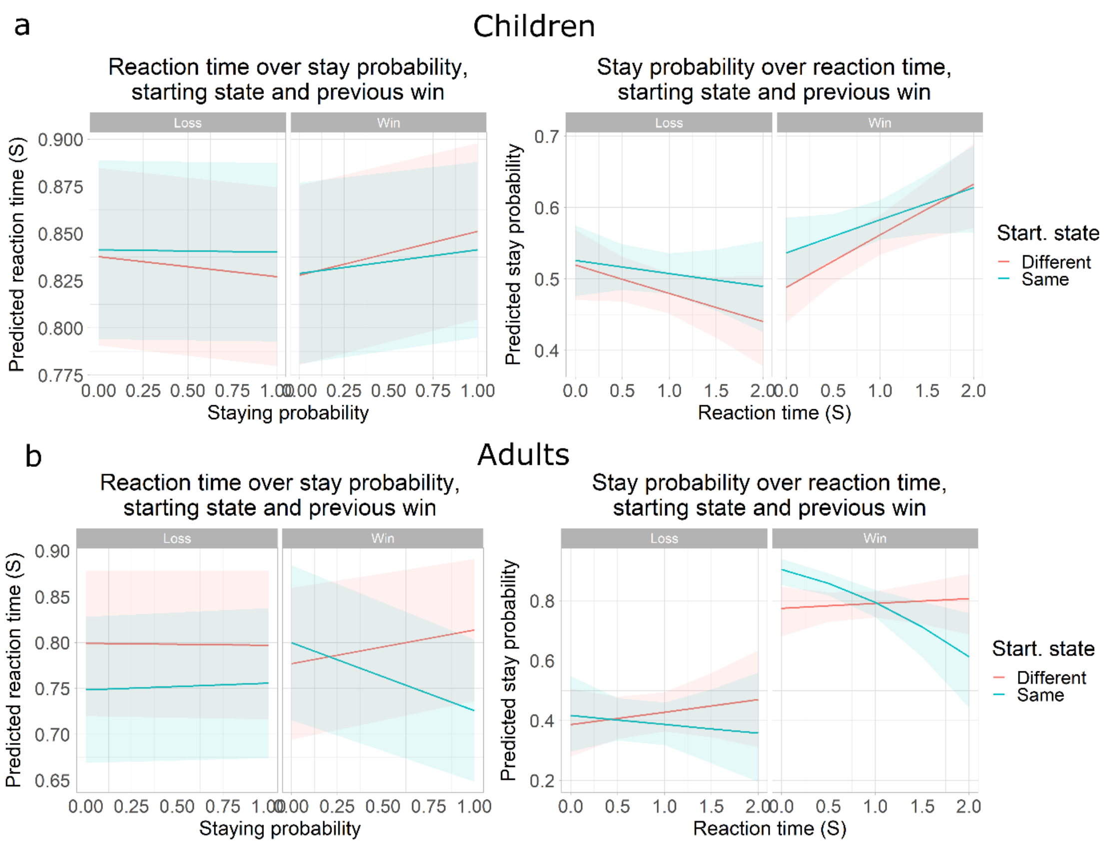


Figure 17. Assessing model-based decision-making via reaction times. There were some indications that reaction times reflected processing time after previous wins and depending on starting state, but these were not the same between children and adults.

2.6.9 Behavioral stay-plots for children and adults

A limitation of the current task is that there is no obvious behavioral pattern of model-based and model-free decision-making. For example, for the participants, there is no obvious measure of a reward that is lower than expected or higher than expected without going back to the model-derived parameters, such as prediction errors. To visualize pure behavioral choices on this task, I have created a binary variable that indicates whether the previous reward won on a trial was more or less than the participant's mean reward rate. This approximates a higher than expected or lower than expected reward, as participants do not observe all rewards of each action as they progress through the trials.

Using this method, I created par plots of behavioral staying (Figure 18). These plots show the behavioral staying percentage as a function of the current starting state, whether the previous trial was a “win” or a “loss” over low and high stakes for adults. For visualization, I used a median split for children showing this for older (N = 43) and younger (N = 42) children in the sample. However, I would like to point out that these bar plots can only be approximations of behavioral model-based decision-making and do not capture the full extent of human behavior and learning on this task. I created these bar plots to illustrate how model-free and model-based agents could act on this task, disregarding any other learning or behavior.

This plot indicates that the children do not adhere to the pattern of a pure model-free behavioral agent; instead, they show the model-based behavioral staying pattern, albeit less strongly than the adults. This is also reflected in adults' propensity to show a higher degree of behavioral staying overall ($F(1,12631) = 120.9$, $p < .001$). For the older children, I see that they more closely reflect the model-free pattern during low-stake trials, while during the high-stake trials, they more closely reflect the model-based pattern. I see the opposite for the younger children in that their behavioral pattern more closely resembles a model-based pattern during the low-stake trials.

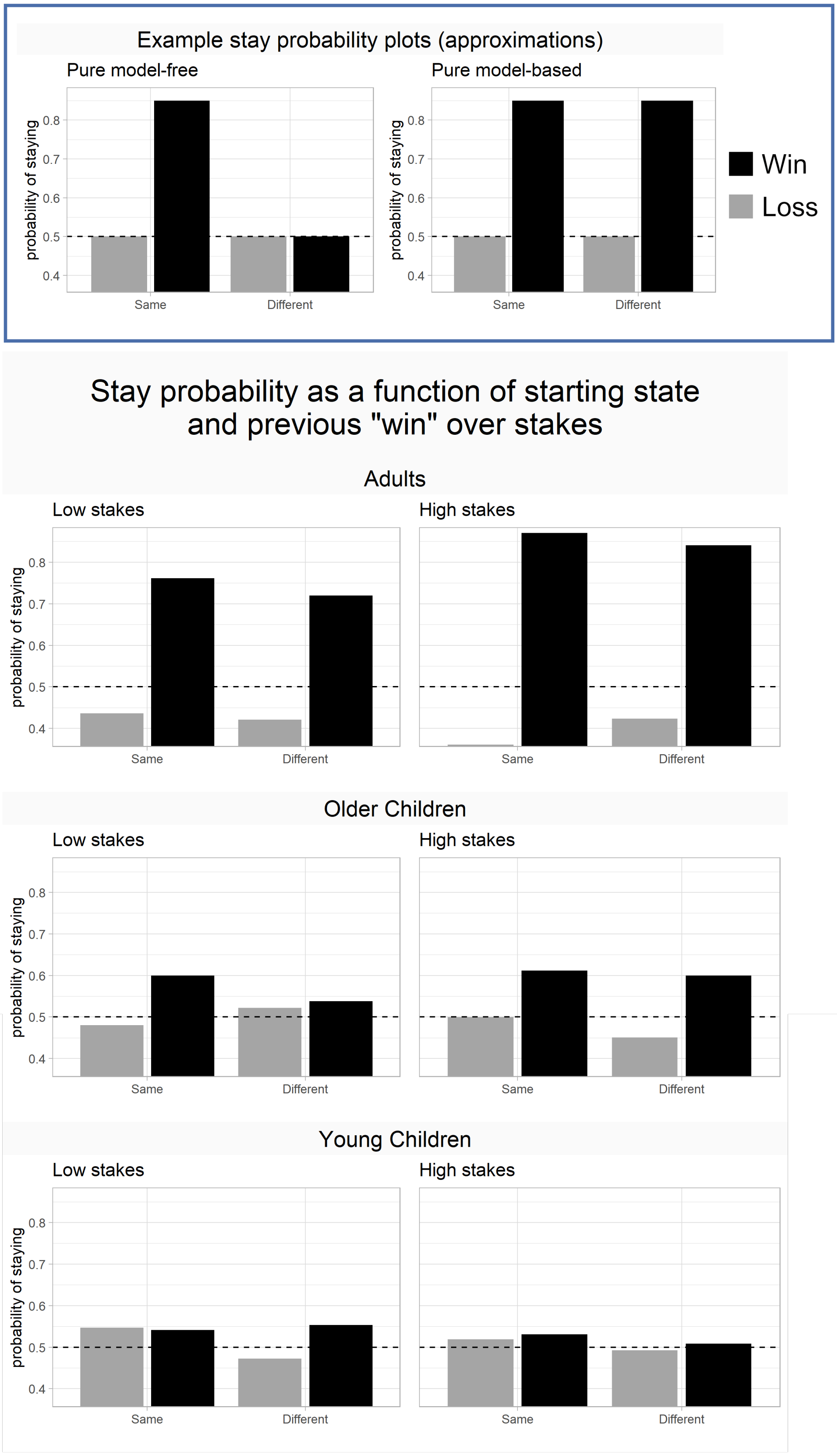


Figure 18. Bar plots of behavioral staying for three age groups and example plots of a pure model-free and model-based agent.

To create a binary behavioral variable of rewards that are lower or higher than expected, I logged whether the previous reward was higher or lower than the participant's mean reward rate. This approximates a higher than expected or lower than anticipated reward. These plots show the behavioral staying percentage as a function of the current starting state, whether the previous trial was a "win" or a "loss" over low and high stakes for adults, and a median split for children showing this for older and younger children in the sample. This plot indicates that the children do not adhere to a pure model-free behavioral agent; instead, they show the model-based behavioral staying pattern, albeit visually less strongly than the adults. This is also reflected in adults' propensity to show a higher degree of behavioral staying ($F(1,12631) = 120.9$, $p < .001$). Older children more closely reflect the model-free pattern during low-stake trials, while during the high-stake trials, they more closely reflect the model-based pattern. For the younger children, the opposite is true, in that their behavioral pattern more closely resembles a model-based pattern during the low stakes.

Chapter 3. The neurocognitive correlates of model-based decision-making and metacontrol in childhood

Part of Chapter 3 is currently under review for publication:

Smid, C. R., Ganesan, K., Thompson, A., Cañigueral, R., Veselic, S., Royer, J., Kool, W., Hauser, T. U., Bernhardt, B & Steinbeis, N. (2022). Neurocognitive basis of model-based decision-making and its metacontrol in childhood.

3.1 Abstract

Human behavior is supported by both goal-directed (model-based) and habitual (model-free) decision-making, each differing in its flexibility, accuracy, and computational cost. The arbitration between habitual and goal-directed systems is thought to be regulated by a process known as metacontrol. However, how these systems emerge and develop remains poorly understood. In the previous chapter, I found that while children between 5-11 years displayed robust signatures of model-based decision-making, which increased during this developmental period, there were substantial individual differences in the display of metacontrol. Here, I inspect the neurocognitive basis of model-based decision-making and metacontrol in childhood and focus this investigation on executive functions, fluid reasoning, and brain structure. A total of 69 participants between the ages of 6-13 completed a two-step decision-making task and an extensive behavioral test battery. A subset of 44 participants also completed a structural magnetic resonance imaging scan. I find that individual differences in metacontrol are specifically associated with performance on an inhibition task and individual differences in dorsolateral prefrontal, temporal,

and superior-parietal cortical thickness. These brain regions likely reflect the involvement of cognitive processes crucial to metacontrol, such as cognitive control and contextual processing.

3.2 Introduction

To engage in optimal decision-making, individuals need to link their actions to associated outcomes. Classical learning paradigms propose that this challenge is solved through the operation of two distinct systems that differ in their flexibility and computational cost, with one operating habitually and the other in a goal-directed fashion (Boureau et al., 2015; Daw, 2018; Daw et al., 2005). Habitual and goal-directed strategies have been formalized in model-free and model-based reinforcement learning algorithms (Daw et al., 2005; Dolan & Dayan, 2013; Gläscher et al., 2010). Model-free decision-making engenders value-based learning, which relies predominantly on tying actions to previous rewards. In contrast, model-based decision-making relies on using an internalized model of the world to match the rewards attained with the appropriate actions depending on the context (Daw et al., 2011; Kool et al., 2016).

Model-free decision-making is not always adequate but is cognitively less costly as it relies on looking at cached values of past actions. On the other hand, model-based decision-making is more accurate and costly, as new values have to be computed continuously (Keramati et al., 2011). Furthermore, optimally responding to different environmental demands, with the inherent processing limits of human cognition, requires dynamic arbitration between the costs and benefits of both decision-making systems (Dubois et al., 2022; Lieder

& Griffiths, 2019). For example, for everyday tasks, the efficiency of habitual decision-making might be preferred and allows saving of cognitive resources, while to be successful in novel or complex scenarios, more goal-directed methods may be required. Human decision-making, therefore, continuously requires the arbitration of the potential rewards and costs associated with each action (Bolenz et al., 2019; Boureau et al., 2015; Ruel et al., 2021), a process known as metacontrol.

Prior work found that the display of model-based decision-making emerged only in adolescence and increased through adulthood when using decision-making tasks originally designed for adults (Decker et al., 2016; Nussenbaum et al., 2020; Palminteri et al., 2016; Potter et al., 2017). Recently, it has been shown that children as young as five displayed model-based decision-making and that its use continuously increased throughout development (Smid et al., 2022). Notably, the dynamic deployment of these model-based vs. model-free systems seems to be a process that only emerges later in life. By manipulating the reward one could gain, I showed that adults dynamically increase their model-based reasoning for bigger rewards in the previous chapter, a process termed metacontrol. In contrast to adults, children did not display optimal metacontrol, as indicated by prioritizing model-based decision-making for high-stake rather than low-stake trials. Instead, children in this age range showed substantial individual differences in metacontrol, rendering it a critical period to better understand the neurocognitive correlates that enable this metacontrol.

Cognitive abilities that encompass our ability to expend attention and focus on a task at hand, or our ability to flexibly shift between different tasks are defined as Executive Functions (EFs) (Diamond, 2013). EFs are broadly defined as functions in the realm of working memory, cognitive flexibility, and cognitive control, or the ability to inhibit prepotent impulses (Diamond, 2013). Correlational evidence and experimental manipulations suggest that working memory and inhibition are relevant to model-based decision-making in adults and may underlie this process (Otto et al., 2015; Otto, Raio, et al., 2013; Potter et al., 2017). Further, in a sample of 9-25-year-olds, it was shown that fluid reasoning was linked to model-based decision-making (Potter et al., 2017). In contrast, the neurocognitive foundations of efficient metacontrol are much less studied (Bolenz et al., 2019; Kool et al., 2017; Kool & Botvinick, 2014). Furthermore, while metacontrol appears to be present during adolescence, increases into adulthood (Bolenz & Eppinger, 2021), and decreases in older age (Bolenz et al., 2019), its cognitive bases are unclear. However, it has been proposed that EFs might be relevant (Davidow et al., 2018; Dezfouli & Balleine, 2013; Keramati et al., 2011, 2016; J. J. Lee & Keramati, 2017; K. J. Miller et al., 2018; Otto, Gershman, et al., 2013).

Prior work on the neural correlates of model-free and model-based decision-making has sought to uncover distinct signatures of associated prediction errors. Some studies suggest distinct regions for model-based decision prediction errors, such as the posterior parietal cortex (O'Doherty et al., 2015), the dorsomedial prefrontal cortex (PFC) (Doll et al., 2015), and the (dorso) lateral prefrontal cortex (DLPFC) in particular (Beierholm et al., 2011; Cremer et al., 2021; Doll et al., 2015; Gläscher et al., 2010; S. W. Lee et al., 2014; Smittenaar et al., 2013), while model-free prediction errors have been mainly localized to the (ventral) striatum (Beierholm et al., 2011; Gläscher et al., 2010; O'Doherty et al., 2015) or the putamen (Doll et al., 2015, but see also Daw et al., 2011; Sanfey & Chang, 2008). A potential causal role of the DLPFC in model-based decision-making was identified via direct manipulation of the DLPFC via TMS, which led to a reduction in model-based decision-making (Smittenaar et al., 2013).

In contrast, few studies have addressed the neural correlates of metacontrol concerning switching between decision-making strategies (S. W. Lee et al., 2014; O'Doherty et al., 2015). For example, O'Doherty et al. suggested that the arbitration between model-free and model-based systems was encoded by bilateral inferior lateral PFC, the right frontopolar cortex, and the rostral anterior cingulate cortex (O'Doherty et al., 2015). Meanwhile, Lee et al. found that the arbitration between habitual and goal-directed systems depended on activity in the bilateral lateral PFC (S. W. Lee et al., 2014). In addition, a study on adolescents found that the selective upregulation of cognitive control for trials with greater reward in contrast to trials with lesser reward was governed by frontostriatal connectivity (Insel et al., 2017). This could lead to a similar relationship in the context of stake-based metacontrol used in the current study. Taken together, findings from these studies suggest that the DLPFC, in particular, may be implicated in both model-based decision-making and its metacontrol. In the current study, I used cortical thickness as a marker of brain structure and linked these to model-based decision-making and metacontrol. To do so, I employed two methods of assessing the potential relationship with cortical thickness; (1) whole-brain analysis and (2) ROI analysis of the bilateral DLPFC to see if age-independent differences in brain anatomy in 6-13-year-old children are related to model-based decision-making and metacontrol.

In sum, this study aimed to investigate the neurocognitive correlates of model-based decision-making and metacontrol in children aged 6-13. I related model-based decision-making and metacontrol to performance on an extensive task battery comprising several domains of EFs (working memory, inhibition, cognitive flexibility) and intelligence. While I found no behavioral or structural relationships with model-based decision-making, metacontrol was significantly related to individual differences in inhibition and whole-brain cortical thickness of the entorhinal cortex, the superior parietal cortex, and the bilateral DLPFC in an ROI analysis.

3.3 Methods

3.3.1 Participants

A total of 69 (35 female) participants, with a mean age of 8.99 years (SD = 1.57), and an age range from 6.19 to 12.61 years, were recruited from 20 schools in the Greater London area. A subset of the total sample also completed an MRI scan. The final MRI sample consisted of 44 (25 female) participants with a mean age of 9.37 years (SD = 1.53) and an age range of 6.19 – 12.61 years. Ethical approval for this study was obtained from UCL's Research ethics committee in compliance with UK national regulations. Consent and assent from both parents and children were obtained for all participants. Participants took part in a more extensive intervention study that included data collection on three separate occasions. The current data set was collected at the first testing time point.

Participants visited the UCL testing facilities for their testing session, where both behavioral and MRI data was collected. Sessions were conducted on the same day and ranged between 3.5-4 hours in duration. Participants took multiple breaks during the session where they were supplied with snacks and drinks and were allowed to take as many additional breaks as necessary. When participants no longer wanted to continue with the session, the session was suspended.

3.3.2 Model-based and model-free measures of decision-making

Participants completed a sequential decision-making task that allowed dissociation of different decision-making strategies (Kool et al., 2016, 2017). This task was adapted for a developmental sample and was previously conducted with children of a similar age range (Smid et al., 2022). The task used in the current chapter is identical to the task used in Chapter 2 (Figure 8), except for two changes. First, the trial size was reduced from 140 to 102 trials to reduce the time participants spent playing the task while still capturing the model-free and model-based parameters robustly, see 2.6.3 Parameter recovery. Second, the decision in the second stage was left out, whereas, in the previous task version, participants had to press the spacebar to collect the reward. In the current task, participants received the reward immediately. The same computational model and model-fitting procedure were used, as described in 2.6.1 Dual-reinforcement learning model and 2.6.2 Model-fitting procedure.

I 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. Understanding of the task structure was high, with 96% of the participants correctly reporting the task structure. Missed trials were excluded from the analysis as participants did not receive rewards on these trials and, therefore, could not learn from them. Similar to Chapter 2, participants were excluded if they missed more than 30% of the trials. On average, children missed only 0.05% of the trials, and the highest percentage of trials missed was 17%. Therefore, no participants were excluded from the analysis.

3.3.3 Cognitive task battery

3.3.3.1 Inhibition

Four measures of inhibition were used, a Stop-Signal Task (SST), a Stroop task, a Flanker Inhibition Task, and an AX-CPT task.

SST. In the SST, participants have to press in response to a visual go-cue as fast as possible (Figure 19a) but withhold a response when a stop-signal appears (Figure 19b) (Matzke et al., 2018). During the task, participants were asked to press the left arrow key when seeing the go-signal (i.e., a honey pot) on the left side of the screen and the down arrow key when the signal appeared on the right side. Ten practice trials were administered before participants completed 80 trials of the main task. Each trial started with the presentation of a fixation cross of 1250ms. On 25% of the trials, a stop signal (i.e., a picture of bees) was presented after the honey pot. Participants were instructed not to press any key if they saw the stop signal. The stop signal delay (SSD) started at 200ms, decreased by 50ms after a successful stop trial, and increased by 50ms after an unsuccessful stop trial. Participants had to respond within 6-seconds, or the trial timed out. To derive a measure of inhibition, the mean Stop-Signal Reaction Time (SSRT) was calculated using the integration method (Verbruggen et al., 2019). This measure was summed within participants and z-scored. The SSRT was inversely coded for this study to mean that a larger score indicates better inhibition (“SSRT”).

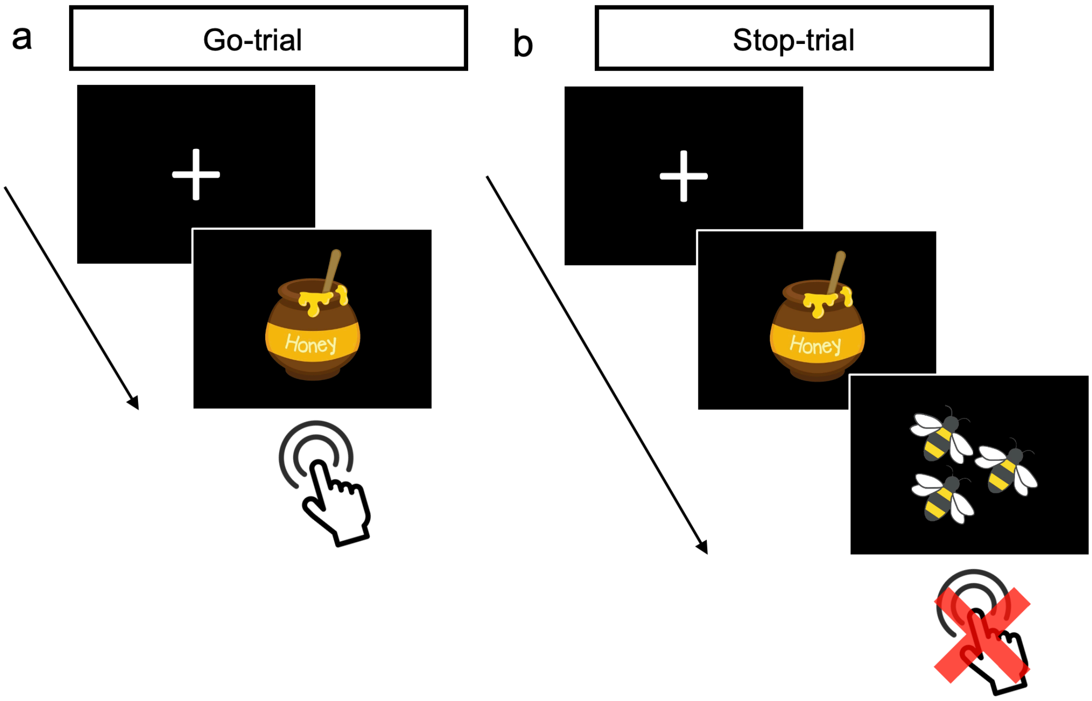


Figure 19. Stop Signal Task (SST).
During a go-trial (a), participants were instructed to react as fast as possible to the go-signal (honey pot) by pressing an arrow key depending on whether the stimulus was depicted on the left or right side of the screen (left and down arrow key). However, during a stop trial (b), the stop-signal was presented after a short stop-signal delay (SSD), and participants were instructed to withhold their response.

Stroop task. Another measure of inhibition was a child-adapted Stroop task, where participants had to respond to congruent and incongruent trials with an auditory cue (Williams et al., 2007). Participants were asked to match animals to where they live (e.g., a frog to a pond). Four animal habitations were presented in the four corners of the screen throughout the game, and participants had to move their mouse pointer to the habitation of the animal on the current trial. At the start of every trial, an animal cartoon was displayed in the center of the screen. Participants were told that sometimes the animals wore disguises and to only respond to an auditory cue indicating the animal type (e.g., frog – *"ribbit"*). On congruent trials, both auditory cues and visual cues matched (e.g., frog presented on screen and *"ribbit"* sound played) (Figure 20a). On incongruent trials, auditory cues and visual cues did not match (i.e., dog presented on screen and *"ribbit"* sound played) (Figure 20b). Participants completed four practice trials, after which they completed 72 trials in the main task, with a 50/50 ratio of congruent and incongruent trials. Participants had to respond within three seconds, or the trial timed out. At the start of each trial, the mouse pointer location was reset to the center of the screen, and participants were presented with a blank screen in the center of the trial for 1000ms. For Stroop performance, the difference between reaction time and error rates was calculated separately for incongruent and congruent trials. Then the reaction time and error rates for each trial type were z-scored and summed. The performance measure used was the difference score between the incongruent minus the congruent trials, where a positive score indicated higher processing costs on the incongruent trials. A lower score indicated less difference in the performance between the incongruent and congruent trials ("Stroop").

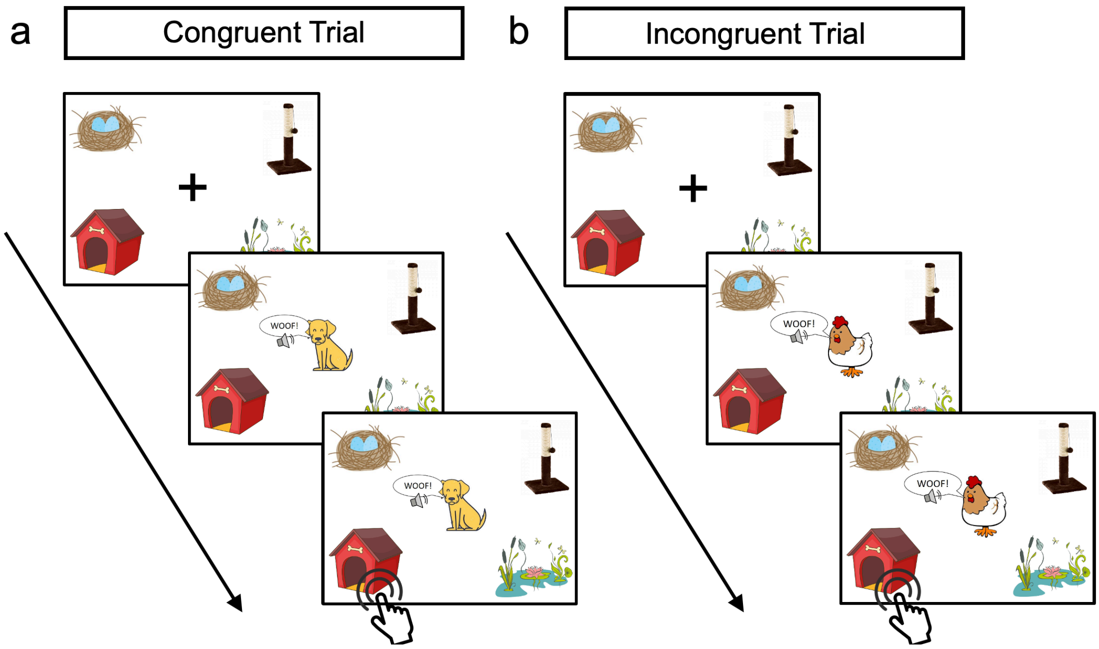


Figure 20. Animal Stroop task.

During the congruent trials (a), the animal depicted in the center and the noise emitted matched, while during the incongruent trials (b), the sound did not match the animal represented. Participants were instructed to ignore the visual center stimulus and respond to the auditory stimulus.

Flanker task (inhibition component). I used an adapted and child-friendly Flanker task with an inhibition component. Participants were shown a row of five fish in the center of the screen for this component. Participants were told to press an arrow key depending on the direction the central visual cue (the middle fish) was facing and to ignore the direction of the distractor stimuli (the flanking four fish). On congruent trials, the central visual goal cue was facing the same direction as the flanking distractor stimuli (Figure 21a), while in incongruent trials, the visual cue was facing the opposite direction from the distractor stimuli (Figure 21b). Participants first completed six practice trials and 40 trials in the inhibition component, with congruent and incongruent trials at a 50/50 ratio. At the start of the trial, participants saw a fixation cross for 500ms, and the central visual cue and the flanking distractor stimuli were shown simultaneously and for 700ms. After this time, the screen became blank, but participants had up to 2.5 seconds afterward to make a response—responses made before 100ms after stimulus onset were not recorded. The ITI was jittered and ranged from 800ms to 2400ms. For Flanker inhibition performance, the difference between reaction time and error rates were calculated separately for incongruent and congruent trials. Then the reaction time and error rates for each trial type were z-scored and summed. The performance measure used was the difference score between the incongruent minus the congruent trials, where a positive score indicated higher processing costs for the incongruent trials. A lower score indicated less difference in the performance between the incongruent and congruent trials (“Flanker_Inhib”).

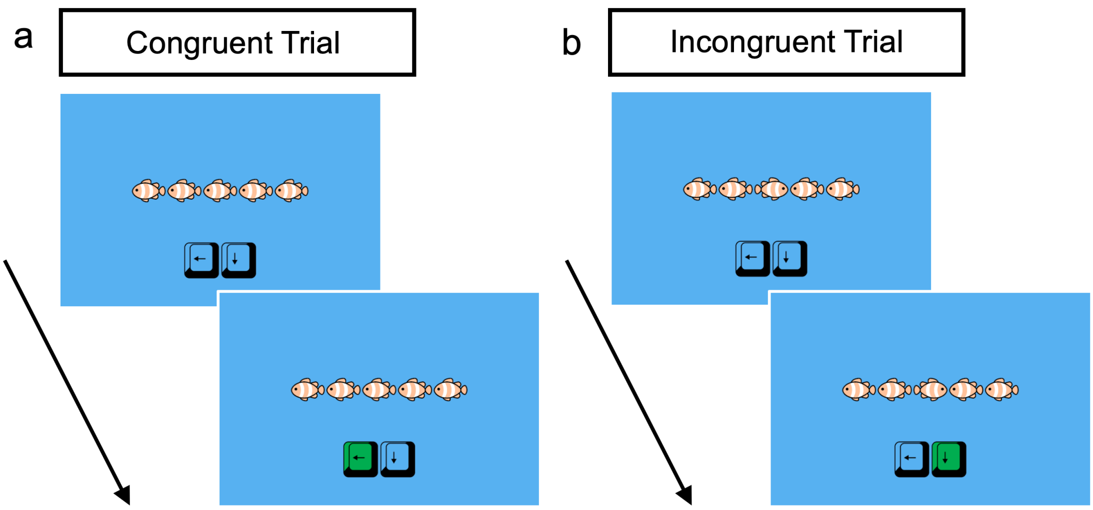


Figure 21. Flanker inhibition task.

During congruent trials (a), the central target stimulus was facing the same direction as the flanking distractor stimuli, while during incongruent trials (b), the central target stimulus was facing the opposite direction from the flanking distractor stimuli. Participants were instructed to always focus on the central target stimulus and respond with a key press in the direction the stimulus was facing (left or down arrow key).

AX-CPT task. Lastly, inhibition was also measured via the AX-CPT task. The AX-CPT task measures participant's tendency to use more reactive or proactive control (Cooper et al., 2017). An A or B cue (i.e., dog or cat) was presented in the middle of the screen for 500ms, followed by an inter-stimulus interval of 750ms and then a probe X or Y stimulus (orange or apple) during which participants had to make their response. Participants had six seconds to respond until the trial timed out. Participants were instructed to press the left arrow key whenever an X followed an A (i.e., AX trials) (Figure 22a) and to press the down arrow key for the presentation of all other cue-probe combinations (Figure 22b). Trials were presented randomly, and 40% of the trials were AX trials, and all other trials (i.e., AY, BX, BY trials) were presented 20% each (Richmond et al., 2015). Participants first completed ten practice trials with feedback, followed by 60 main trials. To measure proactive control. I measured the difference in error rates and response times for the AY trials and the BX trials. I calculated a composite score by deducting the BX trial performance from the AY trial performance and dividing that value by the sum of the AY and BX performance. I then created a composite score by z-scoring these measures and taking the average. When there were zero error rates, these error rates were recoded to 1/2N, where N is the number of trials. This measure is the Proactive Behavioral Index (PBI). It reflects the degree of proactive control displayed during the task, where a higher score reflects more proactive control ("AXCPT") (Gonthier et al., 2016).

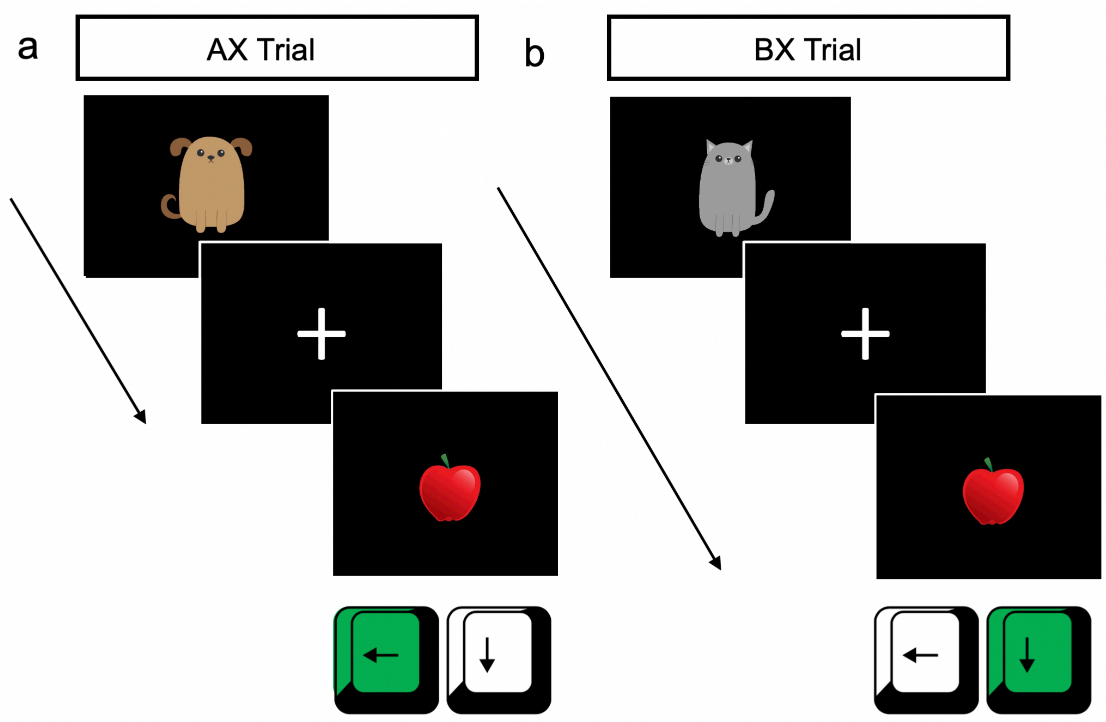


Figure 22. AX-CPT task.
During AX trials (a), participants had to respond by pressing the left arrow key. In contrast, during BX trials (b) (and all other trial combinations, e.g., AY, BY), participants had to respond by pressing the down arrow key.

3.3.3.2 Cognitive flexibility

Two measures of cognitive flexibility were used, the cognitive flexibility task and a Flanker task that measured cognitive flexibility component.

Dimensional switching task. This task assessed participants' ability for rule switching across dimensions (using sound cues (*"animal", "size"*) to respond to either the animal (cat or dog) or the size of the animal (big or small) (Karbach & Kray, 2009). For every trial, a small or big image of a cat or dog was shown in the center of the screen, along with an image of the keys that could be pressed (left and down arrow keys), and the audio cue was played. Underneath each arrow key, the options for the relevant dimensions for that trial were displayed in text (e.g., "small" and "big", or "cat" and "dog"). Participants had 10 seconds to respond before the trial timed out, during which the stimuli remained on the screen—responses made before 200ms after stimulus onset were not recorded. The ITI was jittered and ranged from 1000ms to 1200ms. Stay trials were preceded by a trial in the same dimension (i.e., participants had to respond to the type of animal twice in a row) (Figure 23a). In contrast, during switch trials, the current trial was preceded by a trial in a different dimension (i.e., participants had to first respond to the size of the animal but now to the size) (Figure 23b). Participants completed 20 single-dimension trials in two blocks and 40 mixed trials in one block. They completed separate practice sessions for single and mixed trials with four practice trials where three out of four trials had to be correct to progress. During the single dimension blocks, participants only had to respond to the same dimension (e.g., they only had to respond to the size of the animal), while in the mixed blocks, the two dimensions were mixed. Switch trials were controlled to only occur after either two or three preceding stay trials. Performance on the cognitive flexibility task was captured by the difference in speed and accuracy between the switch and stay trials in the mixed blocks. The reaction times and accuracy for each trial type were z-scored and summed ("CogFlex"). A higher positive score indicated greater processing costs on the switch trials.

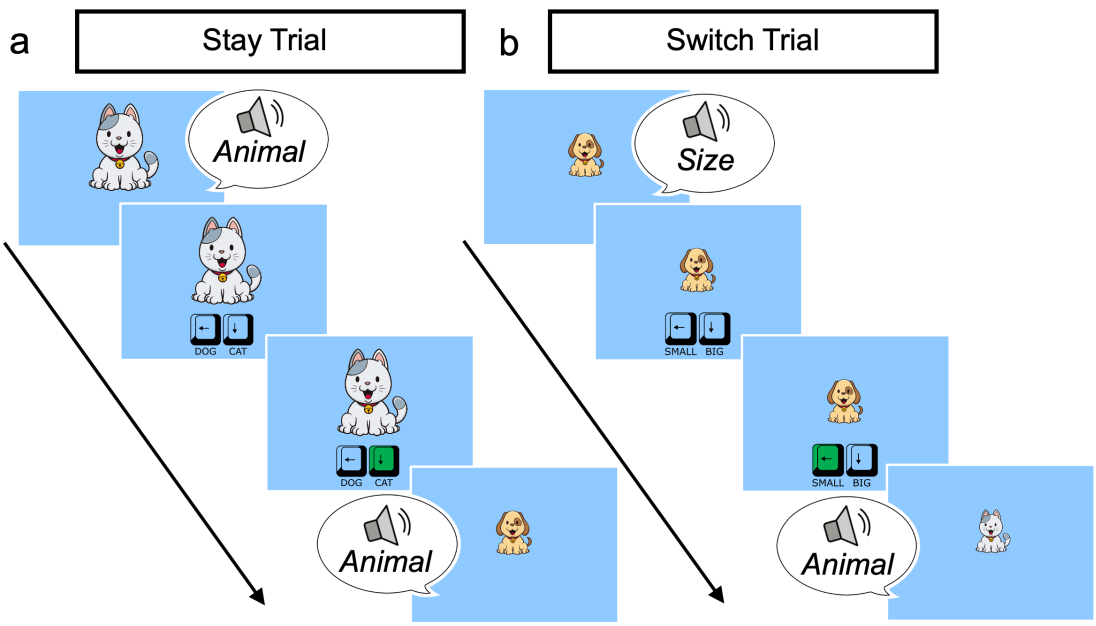


Figure 23. Dimensional switching task.

During stay trials (a), the previous trial was in the same dimension as the current trial, i.e., the participants had to respond to the type of animal displayed and not the size. During switch trials (b), the previous trial was a different dimension than the current trial, i.e., participants had to respond to the size of the animal displayed previously but now have to respond to the type of animal. After a short delay, an image of arrow keys with the current dimension (i.e., animal or size) was displayed under the central target stimulus.

Flanker task (cognitive flexibility component). Participants completed six practice trials before completing 40 trials across two conditions. In the stay condition, the participant had to press the arrow key to match the direction the visual stimuli were facing (the row of five fish, always facing the same direction) (Figure 24a). In the switch condition, as indicated by all five fish changing color, participants had to press in the opposite direction from the way the stimuli were facing (Figure 24b). Stimuli were presented for 700ms, and all responses made before 100ms were not recorded. The ITI was jittered and ranged from 800ms to 2400ms, and participants had 2.5 seconds to respond before the trial timed out. For switching performance, the difference between reaction time and error rates was calculated separately for the switch and stay trials. Then the reaction time and error rates for each trial type were z-scored and summed. The performance measure used was the difference score between the switch minus the stay trials, where a positive score indicated higher processing costs on the switch trials. A lower score indicated less difference in the performance between the switch and stay trials (“Flanker_Switch”).

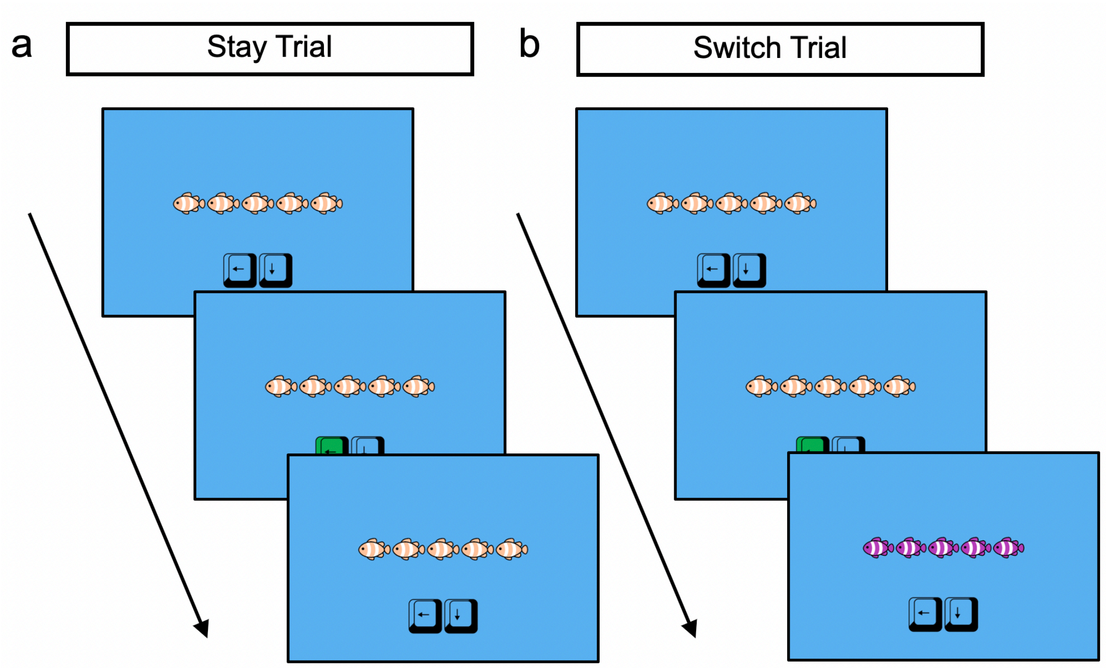


Figure 24. Flanker switching task.
During stay trials (a), participants had to respond by pressing the direction the fishes were facing as fast as possible. In contrast, during switch trials (b), participants had to press the opposite direction from which the fish were facing. In this task, all fish were always facing the same direction.

3.3.3.3 Working memory

Working memory span and manipulation were assessed via two tasks.

CORSI block-tapping task. This task measured visuospatial working memory span with a higher value indicating a higher span (Farrell Pagulayan et al., 2006). This task consisted of a frog jumping between nine potential locations designed as lily pads (Figure 25a). The participants followed the jumps by clicking on the lily pads in a forward sequence (Figure 25b). Participants completed three practice trials with feedback, and two had to be correct to continue to the main task. The main task had 14 trials, and the difficulty changed in a stepwise manner designed as a 1-up, 2-down adaptive staircase. This meant one correct answer added one jump, and two wrong answers removed one jump. Trials commenced with a count-down from three to one, and then the stimulus of the frog jumping was shown for 600ms for every jump. The ISI was fixed to 600ms. The final measure of interest was the highest number of correctly repeated consecutive jumps, referred to as working memory span ("WM_Span").

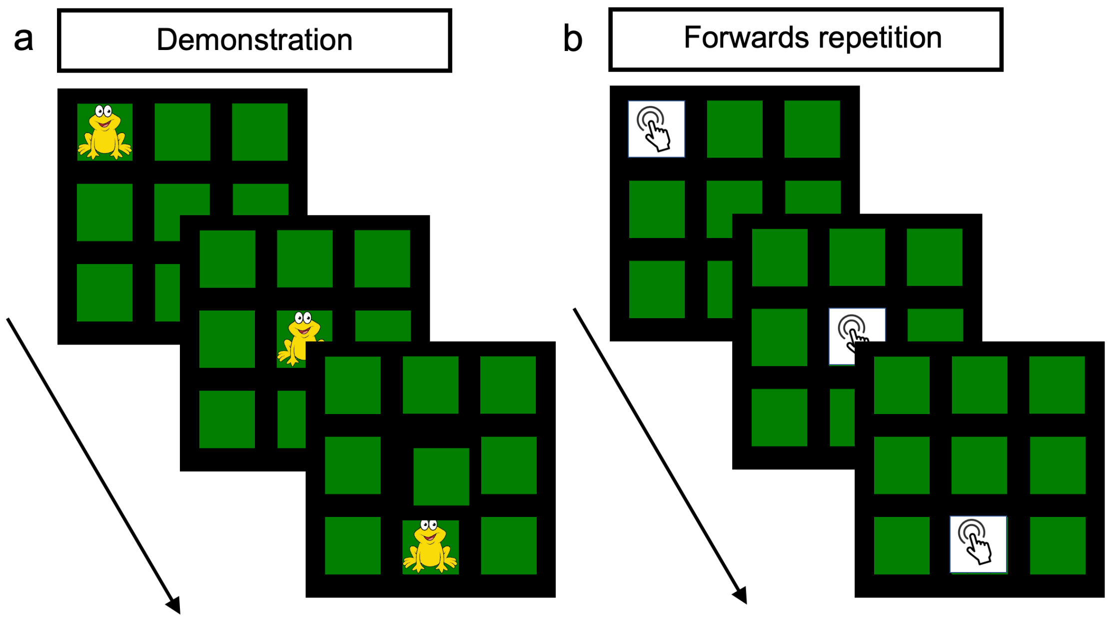


Figure 25. Corsi block tapping task.
For each trial, participants first observed the target stimulus "jumping" between lily pads (a); afterward, participants were required to repeat the forward sequence of jumping by clicking on the corresponding "lily pads" (b).

N-back task. In addition, the n-back task was used to measure working memory manipulation (Chen et al., 2008). For every trial, participants observed a sequence of dinosaurs (center of the screen). In the 1-back condition, participants had to press the spacebar if the current dinosaur on the screen was the same as the previous dinosaur (Figure 26a). In the 2-back condition, participants had to press the spacebar if the current dinosaur on the screen was the same as two dinosaurs previously (Figure 26b). Participants completed 80 trials in total, 40 for each n-back condition. Each dinosaur was shown for 500ms and was followed by a 1500ms Inter-Stimulus-Interval (ISI). Responses made before 100ms after stimulus onset were not recorded, and participants had to make their response before the onset of the next stimulus presentation to be within the response window. The final measures included were the d-prime for both the 1-back and 2-back conditions (“WM_1back”, “WM_2back”).

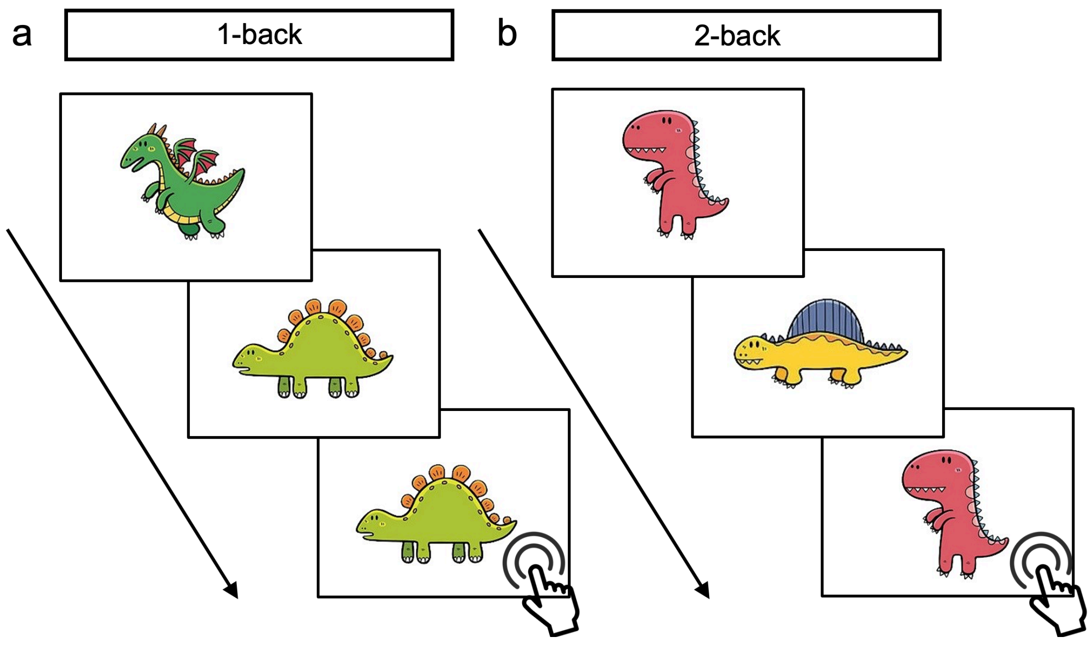


Figure 26. N-back task.

Participants completed two blocks, a 1-back block and a 2-back block. During the 1-back block (a), participants had to respond by pressing the spacebar if they saw the same dinosaur twice in a row. Stimuli were presented sequentially, and only one dinosaur was visible at the time in the center of the screen. During the 2-back block (b), participants had to respond by pressing the spacebar if the current dinosaur was the same as two stimuli previously. Participants were instructed to press as quickly as possible.

3.3.3.4 Intelligence

In addition to the EF tasks, I used two sub-tests of the WASI-II to measure intelligence.

Fluid reasoning. For the fluid reasoning measure, I used the WASI-II Matrix Reasoning subtest (Wechsler, 2011). The Matrix Reasoning subtest was conducted offline and one-on-one by a researcher with a participant and a WASI-II booklet. However, after the Covid-related lockdown, it was administered online via PyschoPy (Peirce, 2007). Participants were asked to choose the image from five options to complete the missing picture in a sequence of images (Figure 27). The task measured pattern recognition, and the correct missing image completed or adhered to the pattern visualized in the sequence of images. The task continued until the participant had three consecutive incorrect answers or until they attained the maximum number of items for their age group. Afterward, their raw scores were converted to standardized scores by age as instructed in the WASI-II manual (“WASI_Matrix”).

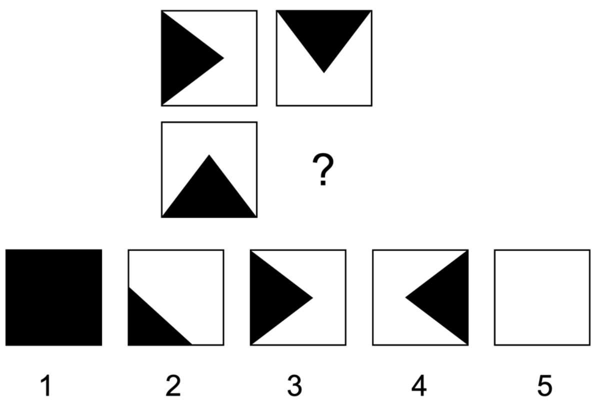


Figure 27. Matrix reasoning example.

Toy example of a matrix reasoning problem. Participants had to complete the sequence by pressing the image that best fits from the five options displayed at the bottom. This example uses a simple rotation rule; the correct answer would be 4.

Crystallized intelligence. The WASI-II vocabulary subtest measured crystallized intelligence (Wechsler, 2011). The Vocabulary subtest was only conducted offline and one-on-one by a researcher with a participant and a WASI-II booklet. This task was not part of the online battery. Participants were asked to describe a word, for example, “shirt”, which had several two-point (e.g., “clothing”), one-point (e.g., “keeps you warm”), and zero-point answers (e.g., “*points at shirt*”) (Figure 28). The task continued until the participant had three consecutive zero-point answers or until they attained the maximum number of items for their age group. Afterward, their raw scores were converted to standardized scores by age as instructed in the WASI-II manual (“WASI_Vocab”).

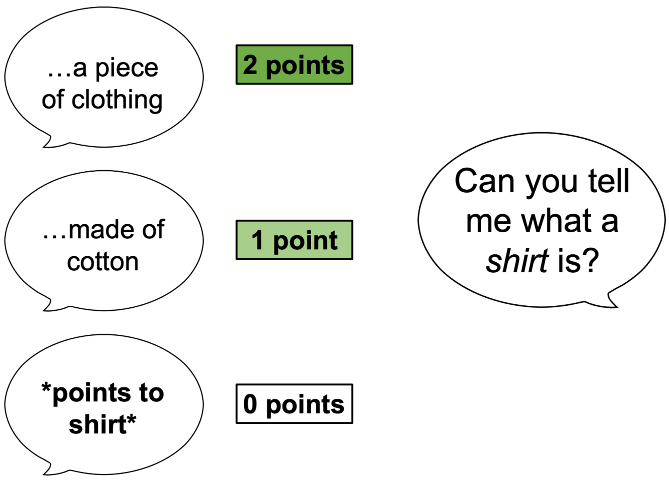


Figure 28. Crystallized intelligence example.
Participants were asked to explain what a word meant. In this case, the prompt was “shirt.” In the text balloons on the left, examples of 2-point, 1-point, and 0-point answers are depicted.

Table 6 reflects the main domains for each task, the task name, and the abbreviation for the final included measure in brackets.

Table 6. Executive function and intelligence tasks

Domain	***Task***	***Main measure and label***
Inhibition	SST	Stop-Signal Reaction Time (SSRT) coded inversely: higher values indicated better inhibitory control (SSRT)
	Stroop	Difference between incongruent and congruent trials in composite scores of speed and accuracy (a higher value indicated greater processing costs during incongruent compared to congruent trials; Stroop)
	Flanker	Difference between incongruent and congruent trials in composite scores of speed and accuracy (a higher value indicated greater processing costs during incongruent compared to congruent trials; Flanker_Inhib)
	AX-CPT	Difference between the AY and BX trials (PBI Index), where a positive value reflected a higher processing cost on AY trials, indicating more proactive control, and a negative value reflected higher processing cost on BX trials, indicating more reactive control; (AX-CPT)
Cognitive Flexibility	Dimension-switching	Difference between switch and stay trials in composite scores of speed and accuracy (a higher value indicated greater processing costs during switch compared to stay trials; CogFlex)

	Flanker	Difference between switch and stay trials in composite scores of speed and accuracy (a higher value indicated greater processing costs during switch compared to stay trials; Flanker_Switch)
Working Memory	Corsi Block tapping	The highest number of correctly repeated consecutive repetitions referred to here as working memory span (WM_Span)
	N-back	Composite scores for both the 1-back and 2-back condition. A higher score indicated better working memory performance for each condition. (WM_1back, WM_2back).
Intelligence	Fluid reasoning	Age-standardized measure of fluid reasoning (WASI_Matrix)
	Crystallized intelligence	Age-standardized measure of crystallized intelligence (WASI_Vocab)

3.3.4 MRI acquisition and cortical thickness measurements

High-resolution T1-weighted images were obtained using a Siemens 3.0 Tesla Prisma scanner located at the Birkbeck-UCL Centre for Neuroimaging (BUCNI) equipped with a 32-channel whole-head coil. Images were acquired using a 3D-TFL pulse sequence with a flip angle of 9; Echo Time was set to 0.00298, and Repetition Time to 2.3. Two hundred eight slices with a voxel size of 1x1x1 mm3 were collected, and the acquisition matrix ranged over 256 x 256. To limit head motion, children were requested to keep their heads as still as possible, and foam inserts were placed between the head and head coil to ensure the head was snug in the coil. Visual stimuli were projected onto a screen in the magnet boar that could be viewed via a mirror attached to the head coil. Participants watched cartoons without sound during the acquisition of the structural scan.

Structural MRI images were processed with FreeSurfer (Version 6.0.0; http://surfer.nmr.mgh.harvard.edu (Fischl et al., 2002), a software that can label and segment cortex and white matter. After converting the Dicom files to Nifti using dcm2niix (X. Li et al., 2016), scans were run through FreeSurfer. Then, all scans were visually inspected for quality, and the segmentation was manually corrected in FreeSurfer if not successful. Together with three other independent validators I analyzed the scans and corrected them if needed. After corrections, scans were re-segmented using FreeSurfer, until, upon visual inspection, the segmentation quality was adequate, or if it did not reach the final level of acceptance, excluded. As the final validator I performed a final inspection of all scans. Using this method, 44 MRI scans were included, while one scan was left out of further analysis due to excessive movement and poor segmentation. Given the extensive and robust evidence of causal involvement of DLPFC in model-based decision-making (Beierholm et al., 2011; Cremer et al., 2021; Doll et al., 2015; Gläscher et al., 2010; S. W. Lee et al., 2014; Smittenaar et al., 2013), region of interest (ROI) analyses focused exclusively on this area. To create a DLPFC ROI, the Desikan-Killiany atlas was used (Desikan et al., 2006).

After preprocessing, sulcal and gyral features across individual subjects were aligned by morphing each subject's brain to an average spherical representation that accurately matches cortical thickness measurements across participants while minimizing metric distortion. For whole-brain analysis, thickness data were smoothed using a 10 mm Gaussian kernel before statistical analysis. Selecting a surface-based kernel reduces measurement noise but preserves the capacity for anatomical localization, as it respects cortical topological features (Bernhardt, Klimecki, et al., 2014; Lerch & Evans, 2005).

Given the extensive and robust evidence of causal involvement of DLPFC in model-based decision-making (Beierholm et al., 2011; Cremer et al., 2021; Doll et al., 2015; Gläscher et al., 2010; Smittenaar et al., 2013; Wan Lee et al., 2014), region of interest (ROI) analyses focused exclusively on this area. To create the Region of Interest (ROI) of the DLPFC, the Desikan-Killiany atlas was used (Desikan et al., 2006). This atlas allows automatic division of the cortex into standard gyral-based neuroanatomical regions. This atlas divides the cortex into 34 cortical ROIs into each of the individual hemispheres. I extracted the individual cortical thickness of the ROI that most closely matches the DLPFC in the Desikan-Killiany atlas (ROIs 28 (left) and 64 (right); the Rostral middle frontal cortex) for the ROI analysis.

Cortical thickness data were analyzed using the SurfStat toolbox for Matlab [https://www.math.mcgill.ca/keith/surfstat, (Worsley et al., 2009)]. Cortex-wide linear models were used to assess the effects of age, sex, model-based decision making, and metacontrol on thickness at each vertex. Findings from the surface-based analyses were controlled for multiple comparisons

(Bernhardt, Klimecki, et al., 2014; Bernhardt, Smallwood, et al., 2014; Steinbeis et al., 2012; Worsley et al., 2009). This reduced the chance of reporting a family-wise error (FWE). The cluster-defining threshold was set to $p < 0.01$ and the FWE to $p < 0.05$ (Bernhardt, Klimecki, et al., 2014; Bernhardt, Smallwood, et al., 2014). Mediation analysis was conducted in Python using the Pingouin package (Vallat, 2018). The data and code to run the analyses reported in this chapter can be found on my Github: https://github.com/ClaireSmid/Neurocognitive_Basis_Metacontrol

3.4 Results

3.4.1 Markers of model-based decision-making and metacontrol

To assess whether children were sufficiently engaged with and able to perform the task, I compared their performance to chance level. Task performance 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 8b). Scores lower than zero indicate performance worse than chance, and scores higher than zero indicate better than chance performance (The same method as used in 2.4.1 Children perform above chance level and are not random). The mean corrected reward for children was significantly higher than chance ($t(68) = 5.10$, $d = .61$, $p < .001$, 95% CIs [.015, .034]) and performance was significantly positively correlated with age ($r = .27$, $p = .023$, 95% CIs [.04, .48]). This suggests that the children were able to perform the task and that performance improved throughout childhood.

Model-based decision-making was positively correlated to age (r = 0.25, p = .036, Figure 29a), while metacontrol was again not significantly correlated to age (r = 0.07, p = .549, Figure 29b). A higher degree of metacontrol was correlated to higher model-based decision-making overall (r = 0.28, p = .020). There was also no stakes effect for children as a group (t(132.81) = -1.14, p = .255, Figure 29c). The findings in the current paper thus replicate the previous computational findings in a new sample in childhood (Smid et al., 2022).

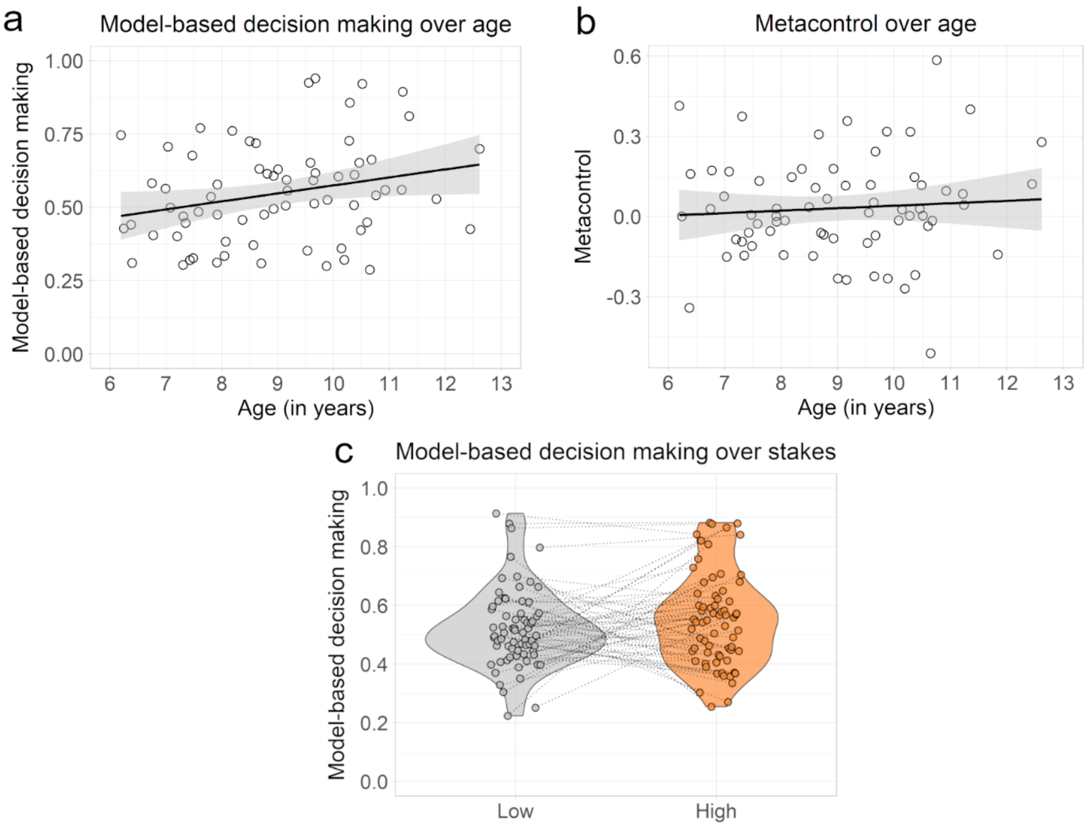


Figure 29. Computational results for model-based decision-making and metacontrol.

(a) model-based decision-making significantly increased with age, (b) while metacontrol did not increase over age, but metacontrol shows substantial individual differences. (c) there was no significant difference in the amount of model-based decision-making over the low- and high-stake trials.

3.4.2 Relationships between model-based decision-making, metacontrol, and executive functions

Next, I assessed the relationships between EFs and model-based decision-making and metacontrol using simple bivariate correlations (Figure 30). Model-based decision-making was positively correlated with working memory span ($r = 0.25$, 95% CI [0.01, 0.46], $p = .039$), indicating that a higher working memory span was correlated to a higher display of model-based decision-making. Model-based decision-making was also negatively related to the cognitive flexibility task-switching measure ($r = -0.30$, 95% CI [-0.50, -0.07], $p = .011$), which indicates that it was related to lower processing costs during the switch trials, see Table 6.

Metacontrol was positively correlated to the Stroop measure ($r = 0.24$, 95% CI [-0.004, 0.45], $p = .046$), indicating higher processing costs on the incongruent trials on the Stroop task may be related to better metacontrol. Metacontrol was positively correlated with the Flanker inhibition measure ($r = 0.42$, 95% CI [0.21, 0.60], $p < .001$), indicating that metacontrol was related to higher processing costs on the incongruent trials.

Significance was corrected for multiple comparisons using Bonferroni (21 tests, threshold at $p = .0023$). Whereas model-based decision-making did not correlate significantly with any measures after correction, metacontrol remained positively correlated with the Flanker Inhibition measure.

To assess this relationship, both the metacontrol and the Flanker inhibition measure were split into their separate components, which for metacontrol consisted of model-based decision-making for the low- and high-stake trials, and for the Flanker measure, processing cost on the congruent and incongruent trials, respectively. Model-based decision-making during high-stake trials was significantly negatively correlated to processing cost on the congruent trials ($r = -0.29$, $p = .014$), while there was no significant relationship with processing cost on the incongruent trials ($r = -.05$, $p = .668$). For model-based decision-making during low-stake trials, there was a significant negative relationship between processing cost during the incongruent trials ($r = -0.24$, $p = .045$), while there was no significant relationship to processing costs on the congruent trials ($r = -.05$, $p = .661$).

Thus, a higher degree of model-based decision-making during the high-stake trials was related to reduced processing cost or improved performance on the congruent trials. On the other hand, a higher degree of model-based decision-making during the low-stake trials was related to reduced processing cost or improved performance on the incongruent trials. Both the metacontrol measure (model-based decision-making during the high-stake trials *minus* model-based decision-making during the low-stake trials) and the Flanker inhibition measure (processing cost during incongruent trials *minus* processing cost during congruent trials) are difference scores. Thus, this suggests that reduced processing cost improved performance on congruent trials is driving the relationship between metacontrol and Flanker.

I also conducted a regression analysis combined with permutation importance testing to determine the best-predicting EFs for model-based decision-making and metacontrol. While model-based decision-making could not effectively be predicted based on EFs, for metacontrol, both Flanker inhibition and Stroop performance were the best predicting features in a non-linear Support Vector Machine regression. As the cortical thickness analysis is based on linear models, these results were not further used in this chapter. For the full regression analysis approach and methods, see 3.6.2 Regression models exploring the relationship between executive functions, model-based decision-making, and metacontrol.

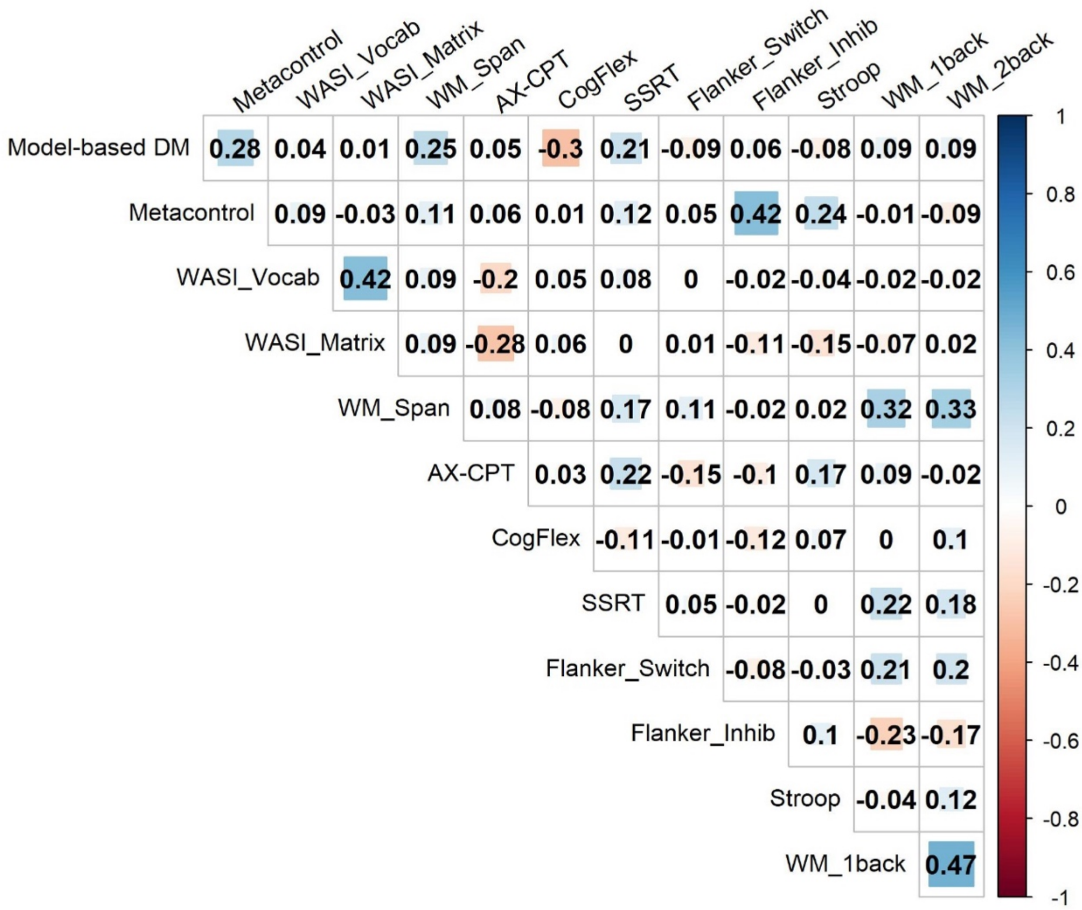


Figure 30. Correlation plot of model-based decision-making, metacontrol, executive functions, and intelligence measures.

For a list of the measures and their acronyms, see Table 6. Executive function and intelligence tasks, and for descriptions of the task, see 3.3.3 Cognitive task battery. The numbers indicate Pearson’s r values.

3.4.3 Cortical thickness, model-based decision-making, and metacontrol

We assessed the relationship between individual differences in model-based decision-making, metacontrol, and cortical thickness. We ran cortex-wide linear models correcting for age and sex and corrected the p-values with FWE and thresholded for significance at $p < 0.01$. We also ran a cortical thickness ROI analysis using the bilateral DLPFC.

No relationship was found between cortical thickness and indices of model-based decision-making at the whole-brain level. For metacontrol, two clusters survived whole-brain correction (Figure 31). Participants with higher metacontrol showed greater cortical thickness in the left temporal lobe encompassing the fusiform gyrus, entorhinal cortex, parahippocampal gyrus, and the right parietal lobe, including the postcentral gyrus and superior parietal cortex, as determined using the Desikan-Killiany atlas (Desikan et al., 2006).

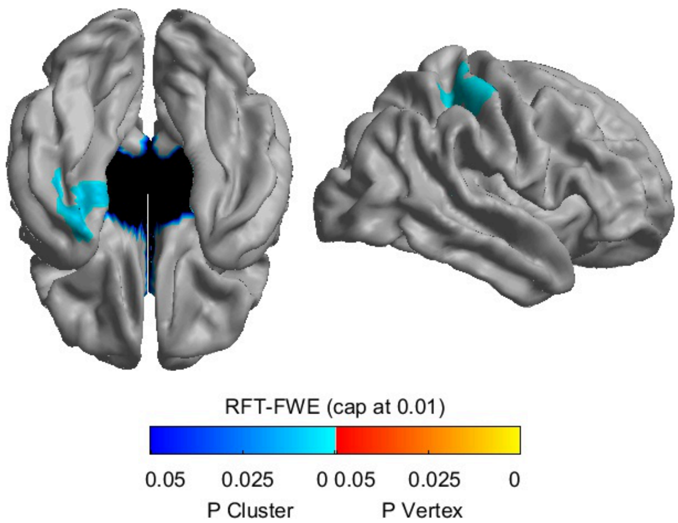


Figure 31. Significant whole-brain clusters of cortical thickness associated with individual differences in metacontrol corrected by age and sex.

Clusters were corrected by age and sex *(thresholded at $p < 0.01$).*

3.4.4 DLPFC ROI analysis

As previous studies have found potential causal links between model-based decision-making and metacontrol and DLPFC (Smittenaar et al., 2013), I also assessed the relationship between cortical thickness in DLPFC bilaterally (Figure 32a). After controlling for age, I ran the ROI analysis with the residual cortical thickness of the DLPFC. As I did not find sex-related differences, I did not control for sex. While I did not find a relationship between thickness in DLPFC and model-based decision-making ($p > 0.09$), metacontrol was significantly related to both cortical thickness in left and right DLPFC ($T(42) = 2.61$, $p = .012$; and $T(42) = 3.00$, $p = .005$ respectively; Figure 32b and Figure 32c). These correlations survived Bonferroni correction (threshold at $p = .0125$). Thus, higher metacontrol was significantly correlated to increased cortical thickness in the bilateral DLPFC for 6–12-year-old children.

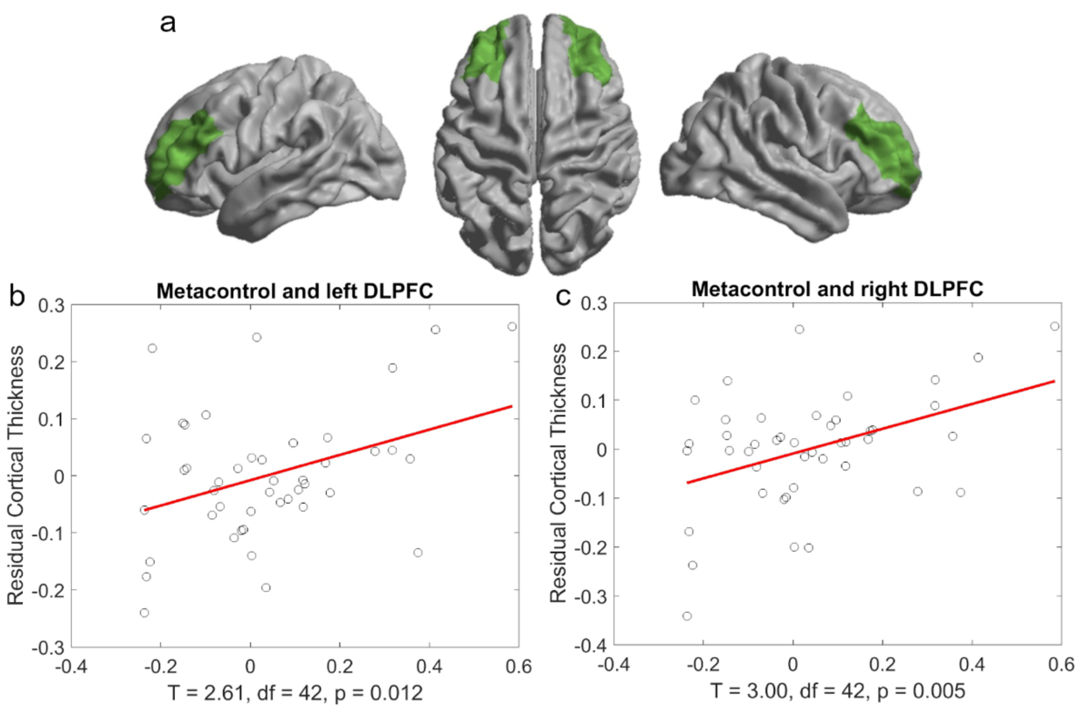


Figure 32. Cortical Thickness of the bilateral PFC and metacontrol.
(a) the ROI for the DLPFC used in the current study is based on the Desikan-Killiany atlas. (b) scatterplot of the relation between metacontrol and residual cortical thickness of the left DLPFC (c) and right DLPFC after correcting for age.

3.4.5 A potentially mediating effect of Flanker inhibition on metacontrol and cortical thickness

Finally, I investigated whether the Flanker Inhibition measure mediated the relationship between the cortical thickness of the bilateral DLPFC and metacontrol. To assess this, I performed a mediation analysis with the Flanker inhibition measure as the potential mediating pathway between cortical thickness and metacontrol. It is important to note that the current sample is small for a mediation analysis, and ideally a much larger developmental sample would have been available. Thus, the results from the mediation analysis should be interpreted with caution. For neither the left (indirect: beta = 0.11, se = 0.09, p = .200, 95% CI [-0.02, 0.24], Figure 33a), nor the right DLPFC (indirect: beta = 0.14, se = 0.13, p = .284, 95% CI [-0.10, 0.43], Figure 33b) was there a mediating effect of inhibition performance on the Flanker task.

I also assessed whether the whole-brain analysis of cortical thickness was improved by adding Flanker Inhibition as a term, but it did not improve the model fit, see 3.6.4 Flanker inhibition's potential effect on whole-brain models of cortical thickness and metacontrol.

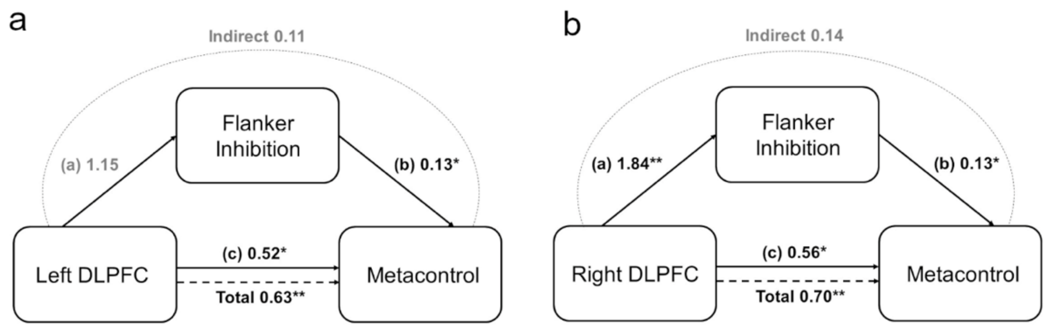


Figure 33. Mediation analysis of the effect of inhibition on the relationship between DLPFC cortical thickness and metacontrol.

(a) Mediation model for the left DLPFC, (b) and the right DLPFC. Cortical thickness entered into the mediation analysis was the residual cortical thickness after correcting for age. Asterisks indicate significance (*p < .05*, p < .01**, p < .001****).

3.5 Discussion

The current study investigates the neurocognitive correlates of model-based decision-making and metacontrol in 6-13-year-old children. To this end, I assessed their relationship with an extensive battery of executive functions (EFs), intelligence and brain structure. While I find that model-based decision-making did not show significant relationships with the EF task battery, intelligence, or cortical thickness measures, metacontrol showed a specific relationship with performance on an inhibition measure and cortical thickness in temporal, superior parietal, and prefrontal brain regions.

I report a relationship between metacontrol and performance on the Flanker task. Specifically, I found that better metacontrol was related to higher processing costs in the incongruent trials than in the congruent trials on the Flanker task. Previous work has shown that the preferential allocation of cognitive resources is in part driven by frontostriatal connectivity (Insel et al., 2017) and that considerations of allocating cognitive effort are, in turn, linked to indices of cognitive control (Kool et al., 2017, 2018; Kool & Botvinick, 2014). However, the current findings appear contradictory, where greater metacontrol is linked to less inhibitory control. Alternatively, children may have preferentially allocated cognitive effort in the measure of inhibition, something they have been shown to do from six years onwards (Chevalier, 2018; Ganesan & Steinbeis, 2021). Thus, participants with higher metacontrol may have prioritized the congruent trials as they are easier (Lieder & Griffiths, 2019; Ruel et al., 2021). So, in the context of making decisions for reward, the trials where more reward can be won are prioritized, and in the absence of increased reward for performance, the neutral congruent trials are prioritized at the expense of the more difficult incongruent trials, which seems to reflect effort avoidance.

In an exploratory whole-brain analysis, I found that individual differences in metacontrol were significantly related to two distinct clusters, one in the left temporal lobe and one in the right superior parietal cortex. The temporal lobe cluster spanned areas involved with memory (Druzgal & D'Esposito, 2001; Jessen et al., 2006; Mion et al., 2010; Rodrigue & Raz, 2004), as well as contextual learning (Aminoff et al., 2007; Coutureau & di Scala, 2009; X. Peng & Burwell, 2021). The superior parietal lobe cluster spanned areas that have previously been linked to working memory (Koenigs et al., 2009), cognitive control (Loose et al., 2017), and planning (Randerath et al., 2017). Thus, these clusters span brain regions previously implicated in cognitive abilities relevant to metacontrol. Contextual-based learning is relevant as metacontrol in the current study represents the ability to increase computationally effortful performance when beneficial selectively. In addition, the previous link between the superior parietal cortex with cognitive control and planning is relevant, as active prioritization of when to employ model-based decision-making across contexts relies on being able to control when to use which decision-making strategy and selectively switching between them based on context. Using an ROI analysis, I found that the cortical thickness of the bilateral DLPFC was positively related to increased metacontrol, a brain region known to be involved in cognitive control and computationally effortful decision-making strategies (Beierholm et al., 2011; Cremer et al., 2021; Doll et al., 2015; Gläscher et al., 2010; S. W. Lee et al., 2014; O'Doherty et al., 2015; Smittenaar et al., 2013). Attempts to integrate behavioral and neural measures to account for metacontrol suggest that these account for distinct portions of variance and constitute unique effects.

Given the previous evidence of a relationship between model-based decision-making and EFs and intelligence (Otto et al., 2015; Otto, Raio, et al., 2013; Potter et al., 2017), the absence of such links in the present sample was surprising. At the very least, this finding suggests that the relationship between model-based decision-making and performance on EF tasks is not straightforward, particularly in the absence of information on how effortful and motivating children might have found the EF tasks. Surprisingly, and similarly to the behavioral analyses, neither whole-brain nor ROI analyses point to any specific relationships with model-based decision-making in the study. Even though I collected similar EF measures used in prior work, reporting significant relationships with model-based decision-making, such as working memory, fluid reasoning, and cognitive control. Differences between previous studies (Otto et al., 2015; Otto, Raio, et al., 2013; Potter et al., 2017) and current findings presumably relate to differences in measures and samples. At a minimum, I conclude that the associations between model-based decision-making and performance on cognitive control tasks may not be robust.

A critical difference between the current and previous studies relates to the task used to measure model-based decision-making. Previous studies relied on the traditional two-step task, which uses stochastic transitions and, compared to the presently used two-step task with a deterministic task structure, was difficult and cognitively more demanding. In essence, it is simpler to employ model-based decision-making on the current task, and a higher degree of model-based decision-making is, in turn, coupled with larger rewards (Kool et al. 2016; 2017). It may well be that prior findings of associations with model-based decision-making and performance on cognitively taxing tasks might be related to task complexity, as opposed to true underlying relationships. It should also be noted that correlating task performance indicates associations at the individual difference level and, not necessarily, whether these processes are used in the context of complex decision-making tasks.

This study has several limitations. While the MRI sample used in the current study (N = 44) is relatively large compared to typical developmental neuroimaging studies, it has been recently suggested that sampling errors could drive significant associations and that robust effects will require sample sizes of the order of hundreds or thousands of participants depending on the phenotype in question (Marek et al., 2022). In the present study, the significant association with cortical thickness in a priori regions of interest lends greater credence to the results. However, it must be noted that all measures are correlational. Future research on model-based decision-making and metacontrol in development may wish to adopt experimental manipulations (i.e., dual-task paradigms) to draw stronger inferences, particularly about the role of EFs.

In sum, the current study investigates the underlying neurocognitive mechanisms of model-based decision-making and metacontrol. I could not replicate previously reported relationships between model-based decision-making and EFs, nor any links with markers of brain structure. However, metacontrol was linked to worse performance on inhibition trials and increased cortical thickness in the temporal and superior parietal lobe and the DLPFC. Metacontrol reflects optimal use of limited cognitive resources, and the current findings suggest that during childhood, this is supported by several brain regions linked to contextual learning and cognitive control. Further, the current results suggest that the relationship between model-based decision-making and other cognitive functions is presumably task-dependent. More extensive investigation with a larger battery of tests, bigger samples, and a better characterization of task-specific associations with goals and effort should illuminate how sophisticated value-based decision-making strategies change during development.

3.6 Supplemental material

3.6.1 Behavioral markers of model-based decision-making

I also investigated the behavioral markers of model-based decision-making and metacontrol. I 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 (Kool et al., 2016; Smid et al., 2022), which is the same method used in the previous chapter. Using this method, the model-based component consists of a main effect of the previous reward on the probability 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 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 (Smid et al., 2022). In addition, I included the difference in available reward across the two planets on the previous trial (a proxy of reward history) and stake (high and low stakes) and age as potential predictors of stay probability. I conducted nested model selection to find the best-fitting model to predict stay probability.

The winning model consisted of previous reward and starting state, age, and stake. There was a significant main effect of previous reward on stay probability (β = .36, se = .03, z = 13.20, $p < .001$), indicating a significant effect of the model-based component in the children's behavior. In addition, there was a main effect of stake, meaning that children were more likely to repeat a visit to the same planet for high-stake trials (β = .09, se = .03, z = 3.34, p = .001). There was a significant interaction between previous reward and age, mirroring the computational finding that with age, children showed more influence of model-based decision-making (β = .18, se = .03, z = 6.43, $p < .001$). There was a significant interaction between previous reward and stake, indicating that for the behavioral marker, there did seem to be more model-based decision-making for higher stake trials (β = .06, se = .03, z = 2.26, p = .024). Lastly, there was a significant interaction between stake and age, indicating that with increasing age, children were more likely to repeat a visit to the same planet for high-stake trials (β = .07, se = .03, z = 2.72, p = .007).

Thus, both computational and behavioral markers indicate that overall model-based decision-making seems to increase with age. Via computational makers, there was no group effect of metacontrol nor an increase with age. However, using behavioral measures, I observed markers of metacontrol in the children's behavior, which increased with age.

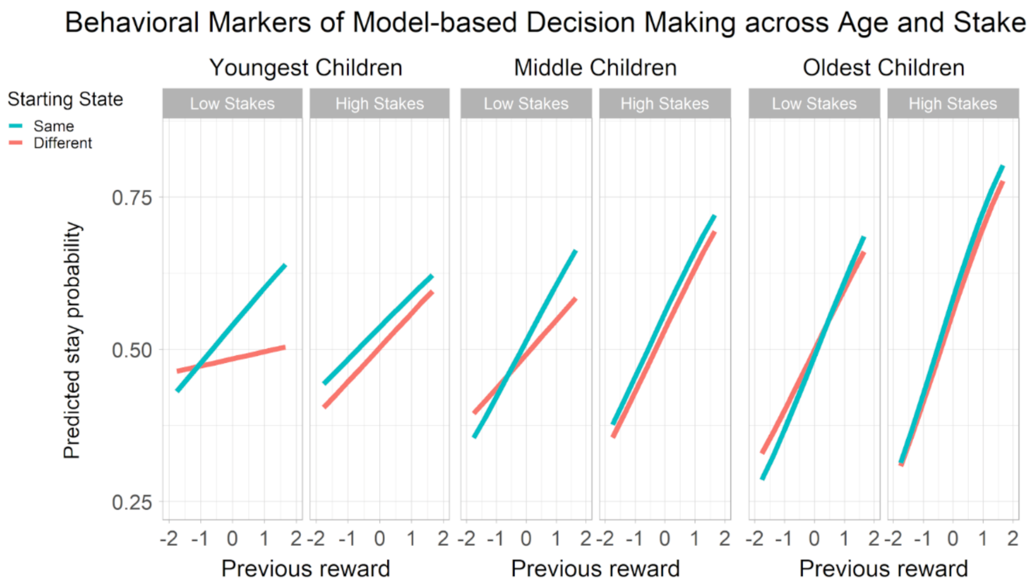


Figure 34. Behavioral markers of model-based decision-making via regression.

The model-based component is reflected in a positive relation between previous reward and stay probability, regardless of starting state. The predicted stay probability is plotted over previous reward, low and high-stake trials, starting state similarity, and age.

3.6.2 Regression models exploring the relationship between executive functions, model-based decision-making, and metacontrol

I also ran a series of regression models to investigate if a combination of EF tasks could predict either model-based decision-making or metacontrol. I included both linear and non-linear models to find the best models and the best predicting EF measures for model-based decision-making and metacontrol. I used a machine learning approach to regression in Python. Regression models were run using the sklearn (Pedregosa et al., 2011) and eli5 packages.

Five regression models were tested (Multiple Linear Regression, Bayesian Ridge Regression, Support Vector Machine (SVM) Regression, Decision Tree, and Random Forest Regression). For each regression model, permutation importance was assessed to rank the best performing EF and intelligence features to predict model-based decision-making and metacontrol. Permutation importance was evaluated in a repeated k-fold cross-validation, using six-folds and 100 repetitions. After finding the best performing features, the hyper-parameters of each regression model were tuned with Leave-One-Out Cross-Validation via grid search. For the k-fold cross-validation and the grid search, the variable to optimize was the negative mean squared error (MSE). The best hyper-parameters were combined with the best predicting features to create the winning model. The performance of the winning model was then assessed via MSE, R-squared (R2), and explained variance, which were obtained in a final k-fold cross-validation. See Figure 35 for a pipeline overview.

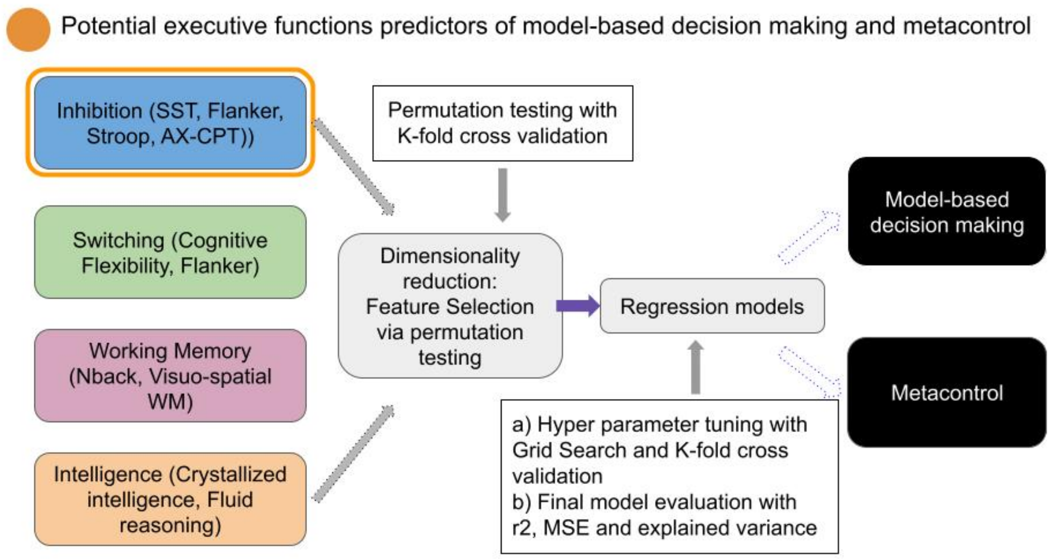
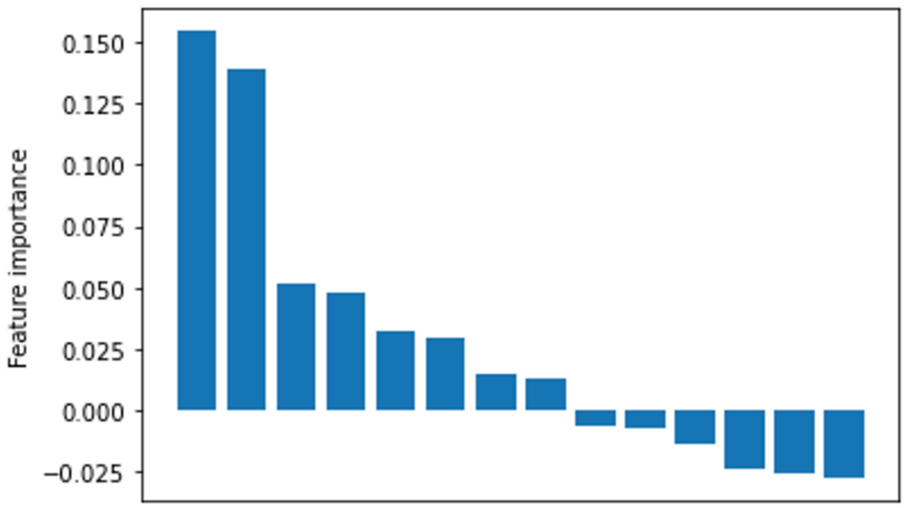


Figure 35. Regression model approach.

None of the regression models reached adequate fit for the assessment of model-based decision-making based on their MSE and R-squared values. This suggests that model-based decision-making was not predicted by the measures of EFs included in this study.

Next, I assessed whether any EF measures were predictive of metacontrol in the sample. The best performing regression model was a Support Vector Machine (Radial Basis Function (RBF) (Gaussian) kernel, C = 10, gamma = 0.01, epsilon = 0.1; MSE = 0.030, R^2 = 0.06, explained variance = 16%), with two EF predictors from the inhibition domain, namely from the Flanker and Stroop tasks.



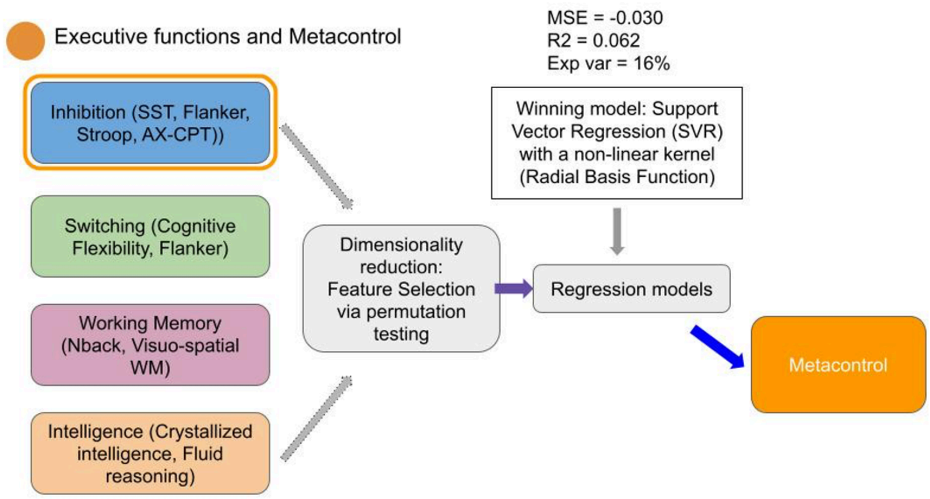


Figure 36. Winning Support Vector Machine (SVM) regression model for metacontrol.

The winning SVM model used a non-linear kernel and two EF measures from the inhibition domain. Feature importance (top) for the included EFs as determined via Permutation testing; each bar represents one measure. Here, the first bar represents The Flanker inhibition, and Stroop measure which were the most predictive features, respectively. Pipeline (bottom) for the winning model and model performance estimates.

3.6.3 Cortical thickness age analyses

Overall mean thickness significantly decreased with age for the sample ($T(42) = -2.34$, $p = .024$), showing that older children had overall thinner cortical thickness. There was no significant difference in the mean cortical thickness between male and female participants ($F1,42) = .21$, $p = .647$).

To investigate whole-brain age-related effects on cortical thickness, I ran a principal component analysis (PCA) to investigate overall structural connectivity across the sample. A PCA of cortical thickness can provide insight into patterns of variations and covariation in the thickness of the cortex across participants in the sample (Yoon et al., 2007). PCA is a statistical technique that reduces the dimensionality of the data by identifying a set of orthogonal, or uncorrelated, components that explain the maximum amount of variation in the data, and each component represents a linear combination of the vertices that contribute to that pattern (Yoon et al., 2007). Thus, the vertices are connected in the PCA based on their similarity in thickness patterns.

The first component explained 12.9% of the variability in the data, the second component explained 5.8%, and the third component 4.8%. Thus, there were no majorly significant principal components to explain the variability in the data. There was a significant relationship between the first principal component and age ($t(43) = -2.23$, $p = .031$), which decreased with age. However, there was no difference in the first principal component and gender ($F(1,43) = .18$, p

= .670). Thus, the first principal component likely reflects a developmental effect in the sample.

The second principal component explained 5.8% of the variability in the data, and this component was also significantly correlated to age. Similarly, to the first component, the strength of the component decreased with age ($t(43) = -2.52$, $p = .015$), thus also reflecting a developmental effect. Likewise, there was no difference in the decrease of strength of this component between the genders ($F(1,43) = .44$, $p = .510$).

Thus, while I observed changes in cortical thickness with age in the sample, I saw fewer differences across the sexes in the current sample.

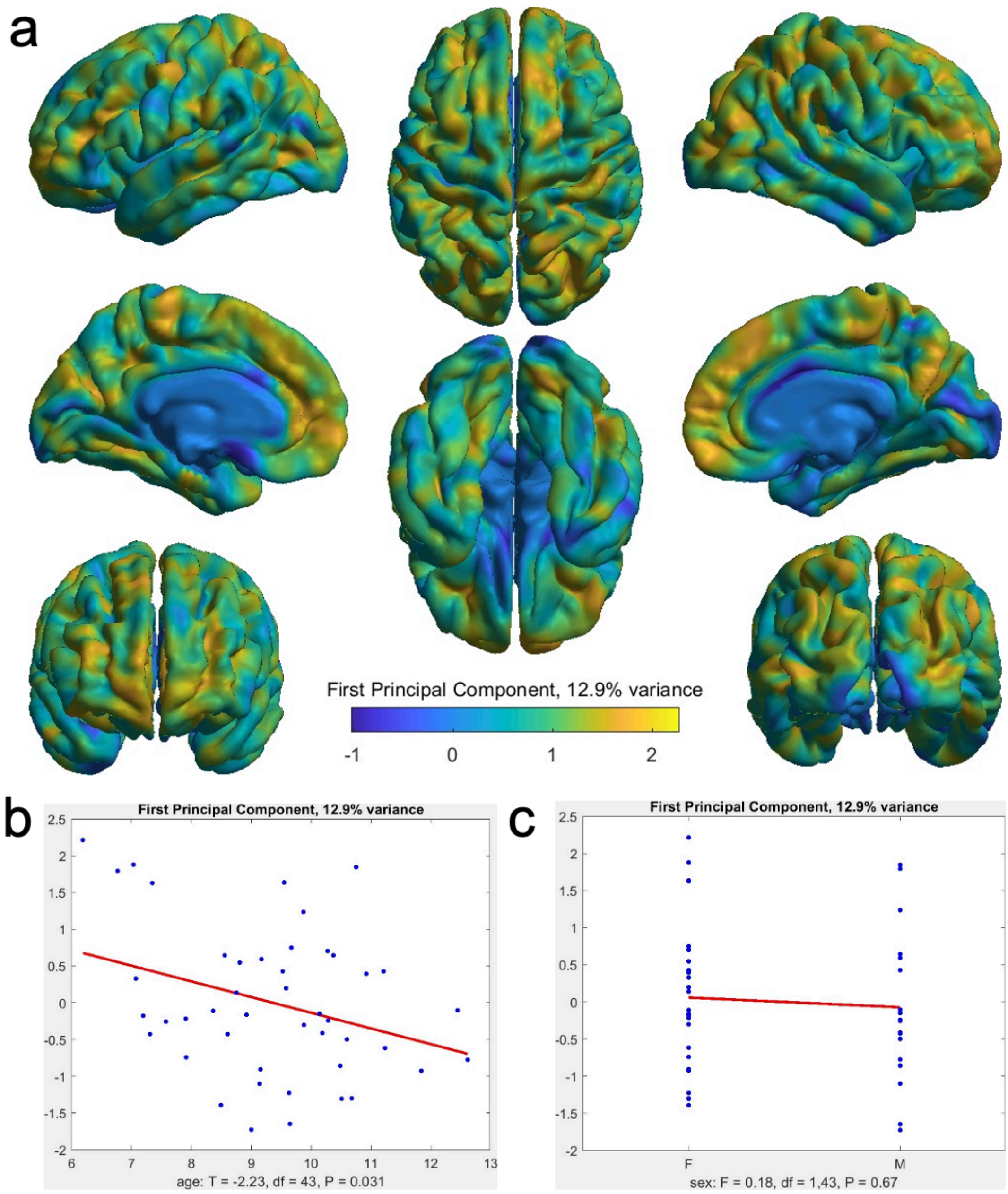


Figure 37. The first principal component of cortical thickness.

Variability of the first principal component for whole-brain cortical thickness (a) and showing the changes in the first principal component over age (b) and sex (c).

3.6.4 Flanker inhibition's potential effect on whole-brain models of cortical thickness and metacontrol

Since I found a potential relationship between inhibition and metacontrol, I sought to investigate how the inhibition measures related to cortical thickness and whether the relationship between cortical thickness and metacontrol would hold or disappear when controlling for the effect of inhibition. If it holds, this suggests that metacontrol might have a distinct relationship to cortical thickness, but if the relationship does not, it suggests that inhibition and metacontrol have shared variance and both influence cortical thickness similarly.

When entering performance on the Flanker task as a term in the whole brain model testing the unique effect of metacontrol while simultaneously controlling for age and sex, all previously reported effects remained significant (Figure 38). In addition, I compared a model with metacontrol, controlled for by age and sex (main model), and a model with metacontrol, controlled for by age, sex, and inhibition (Flanker model), and found that the first model had a numerically better fit (MSE main model = 0.0683, MSE Flanker model = 0.0685, RMSE main model = 0.1367, RMSE Flanker model = 0.1370). The difference in the error sum of squares between the models was significant ($t(20483) = 84.09$, 95% CI [0.060, 0.063], $p < .001$). Thus, adding the Flanker term did not improve the accuracy of the model investigating the effect of metacontrol on whole-brain cortical thickness, and the previously found effects for metacontrol remained significant.

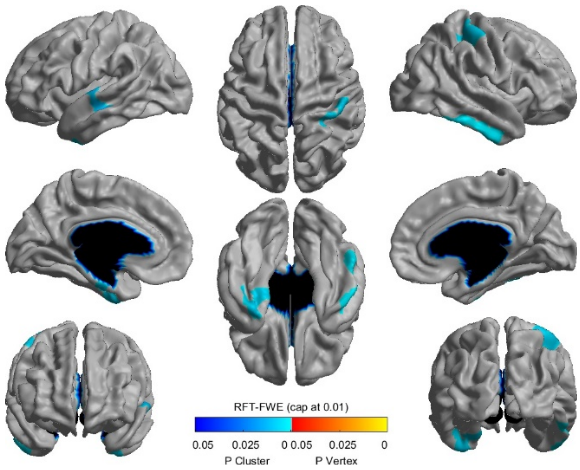


Figure 38. Clusters from a linear model assessing the unique effect of metacontrol on whole-brain cortical thickness, controlling for age, sex, and inhibition.

Chapter 4. Cognitive control and social and intertemporal decision-making in middle childhood

Part of Chapter 4 is currently under preparation for publication:

Smid, C. R., Keertana, G., Thomspon, A., Canigueral, R. & Steinbeis, N. (2022). No effects of cognitive control training on decision-making in middle childhood.

4.1 Abstract

Daily decisions have long-lasting implications for our health, social relationships, and ability to reach long-term goals. In childhood, it is typically observed that with age, decisions become more pro-social in their decisions and discount rewards less with time, indicating that this developmental period is marked by plasticity in decision-making. Although, substantial individual differences in these decision-making abilities are observed independently from pure age-related effects. Alongside improving with age, more pro-social decisions and less temporal discounting are often linked to better executive functions (EFs). EFs are a set of cognitive abilities that encompass working memory, cognitive flexibility, and cognitive control. Thus, individual and non-age-related changes in pro-social and intertemporal decision-making may be reflected in accompanying individual differences in EFs. Furthermore, efforts to enhance EFs via training paradigms have found that these abilities can be improved via training. Thus, improving EFs via a training paradigm in childhood may translate to more pro-social decision-making and less steep temporal discounting. In this study, I investigate (i) the relation between EFs and pro-social and intertemporal decision making in childhood and (ii) whether training

cognitive control leads to changes in pro-social and intertemporal decision-making, using a randomized control trial with a highly variable, adaptive, and complex gamified training protocol in a highly powered sample of 205 children aged 6-13 years. I find that EFs are not strongly linked to pro-social and intertemporal decision-making in this age range and that cognitive control training did not lead to short- or long-term training-related changes.

4.2 Introduction

Our decisions can have far-reaching consequences, as they embody how we interact with others and the world around us. Daily decisions entail different scenarios, from choosing our favorite breakfast cereal to a romantic partner. Typically, decision-making is studied within broad domains related to (i) time (i.e., choosing between rewards of differing magnitude and timescales) (Bickel et al., 2014; Bos & McClure, 2013; Chapman, 1996; Peters & Büchel, 2011; Story et al., 2014), (ii) risk (i.e., choosing between rewards of differing magnitude and likelihood) (Donati et al., 2014; Kacelnik, 1997; S. Li, 2003; Machina, 1987; Verbruggen et al., 2012), and (iii) social contexts (i.e., choosing between rewards for ourself and others) (Böckler et al., 2016; Lockwood & Wittmann, 2018; Lucas et al., 2008). Crucially, our daily decisions can impact real-life long-term outcomes, such as health, wealth, and academic and workplace success (Daugherty & Brase, 2010; Hamilton & Potenza, 2012; Mischel et al., 1988; Silva Castillo, 2017; Story et al., 2014; Tate, 2015; Thaler & Benartzi, 2004). When we turn to developmental samples, we see that with age, children's decisions become progressively more patient (Green et al., 1999; Prencipe et al., 2011; Steinbeis et al., 2016) and more pro-social (Bauer et al., 2014; Chajes et al., 2022; Fehr et al., 2008; McAuliffe et al., 2017), however, at young ages, children tend to make more selfish and impatient choices.

Given the importance of prudent decisions, particularly during development, understanding their underlying mechanisms and seeking to leverage that information to increase it has been a nascent research agenda (Kable et al., 2017; Steinbeis, 2016; Steinbeis et al., 2012; Steinbeis & Over, 2017; Zhao et al., 2022). Childhood is a period of more remarkable plasticity, and if the underlying mechanisms are known and understood, this makes it possible for training paradigms to intervene and boost desired types of decision-making potentially. Here, in a large sample of 6–13-year-old children, the link between decisions related to time and social context (henceforth intertemporal and social decision-making) and executive functions (EFs) are investigated. First, I examine associations between decision-making and latent factors based on an extensive battery of tasks measuring different facets of EFs (i.e., inhibition, working memory, and cognitive flexibility) (see 3.3.3 Cognitive task battery for an overview of the task battery), followed by a considerably sized, randomized control training of inhibitory control or response speed, and studying its short- and long-term effects on decision-making.

Intertemporal choice is deciding between a small reward now or a large reward later; social decision-making is deciding between outcomes for the self or others. EFs reflect a set of processes (e.g., working memory, cognitive flexibility, and inhibition), critical for flexible goal-directed behavior, as they reflect the ability to hold information in the mind and suppress undesired impulses (Diamond, 2013). Childhood EFs have been shown to predict a range of social, academic, and mental health outcomes later in life (Blair & Razza, 2007; Clark et al., 2010; Moffitt et al., 2011). In particular, better EFs have been associated with better personal finances and physical and mental health (Blair & Razza, 2007; Clark et al., 2010; Moffitt et al., 2011). In turn, impaired EF ability has been linked to later mental health problems and criminality in adulthood (Moffitt et al., 2011). EFs undergo protracted development from childhood into early adulthood (Davidson et al., 2006; Garon et al., 2014; Wiebe & Karbach, 2017), which is likely supported by developmental changes in frontoparietal and frontostriatal neural circuitry (Buss & Spencer, 2018; Fiske & Holmboe, 2019). Given their critical role in healthy and productive development and coupled with prolonged plasticity of underlying neural circuitry, EFs have been a primary target for interventions during development (Diamond & Lee, 2011; Heckman, 2006; Johann & Karbach, 2020; Karbach et al., 2015; Karbach & Kray, 2009; Klingberg, 2005; Sala & Gobet, 2017; Wass et al., 2012; Zhao et al., 2022).

EFs are likely crucial in pro-social and intertemporal decision-making, as it allows us to forego our immediate gratification for greater returns, either with time or by investing resources in others rather than ourselves. For example, altruistic behavior entails incurring a cost to oneself (Steinbeis et al., 2012). As a result, pro-social decision-making research has been dominated by a debate between two competing explanations. Some researchers claim that altruistic decisions in adults occur automatically, intuitively, and effortlessly (Rand et al., 2012; Zaki & Mitchell, 2013), while others argue that pro-social decisions require EFs, especially inhibitory control, in the form of self-restraint (Knoch et al., 2006; Rachlin, 2002). Since evidence supporting both views has been observed, findings from developmental psychology may be able to weigh in on this debate under the assumption that developmental origins of altruism may illuminate their presence later in life (Rand et al., 2012; Zaki & Mitchell, 2013). Regarding intertemporal decision-making, young children are particularly prone to making short-sighted choices (Thompson et al., 1997), but with age, they become able to discount rewards less steeply with time (Green et al., 1999; Prencipe et al., 2011; Steinbeis et al., 2016). In previous studies, better cognitive control has been linked to less steep discounting in intertemporal decision-making (Dalley et al., 2011; Figner et al., 2010; Sasse et al., 2017; Steinbeis et al., 2016). Thus, one's EFs may also dictate one's display of pro-social and intertemporal decision-making.

Previous research into cognitive control training has primarily focused on working memory, cognitive flexibility, and to a lesser extent, inhibition (Alloway & Alloway, 2009; Calderon & Newburger, 2018; Holmes et al., 2019; Johann & Karbach, 2020; Karbach et al., 2015; Karbach & Kray, 2009; Karbach & Verhaeghen, 2014; Klingberg, 2005; Zhao et al., 2022), and found that these functions are trainable, although usually in a relatively narrow and task-specific manner (i.e., near transfer) (Diamond & Ling, 2016; Holmes et al., 2019; Kable et al., 2017; Simons et al., 2016). Changes in other distally related domains of cognitive functioning and real-world outcomes (i.e., far transfer) have been much less consistently observed (Holmes et al., 2019; Judd & Klingberg, 2021; Kable et al., 2017; Karbach & Verhaeghen, 2014; Kassai et al., 2019; Sala & Gobet, 2016, 2017, 2019; Scionti et al., 2020; Smithers et al., 2018). While opinions range in their optimism as to the potential for cognitive training to lead to far transfer (Gobet & Sala, 2022; Könen & Karbach, 2015; Sala & Gobet, 2017, 2019), the quality of both training paradigms(Gobet & Sala, 2022; Moreau & Conway, 2014; Raviv et al., 2022; Shawn Green et al., 2019; Smid et al., 2020) and resulting evidence either for or against training-related changes (Dienes, 2014; Dougherty et al., 2016) have been consistently questioned.

There have been previous attempts to influence behavior via cognitive control training paradigms. For example, training paradigms targeting inhibition have been utilized to promote healthier eating and drinking behavior after coupling alcohol or fatty foods stimuli with inhibitory responses (Jones et al., 2016). Given the above evidence on a clear mechanistic link between pro-social and intertemporal decisions and cognitive control, enhancing cognitive control would recommend itself as a candidate to impact decision-making positively. For example, as reduced temporal discounting is potentially supported by working memory, a previous study used working memory training in adolescents to reduce temporal discounting (Zhao et al., 2022). This study found that a working memory training paradigm focused on working memory updating resulted in less steep temporal discounting.

However, another study using a cognitive control paradigm in adults saw no changes in temporal discounting following training (Kable et al., 2017). In addition, as pro-social decision-making has been linked to cognitive control, a previous study sought to improve this via priming inhibition (Steinbeis & Over, 2017). This study found that promoting behavioral control via storytelling increased pro-social decision-making in children. Thus, these studies provide conflicting evidence that enhancements in the executive function domain may transfer to behavioral changes in decision-making tasks. Both studies that observed training-based changes were conducted with developmental samples; perhaps younger ages are more susceptible to training-induced transfer to decision-making tasks (Wass et al., 2012). With previous studies suggesting that the cognitive control mechanism may underpin both intertemporal and pro-social decision-making, improving this mechanism in childhood via training may lead to increases in pro-social and intertemporal decision-making (Kable et al., 2017; Steinbeis & Over, 2017).

Training studies have generally been criticized for lack of effectiveness and transferability (Sala & Gobet, 2019). A few critical issues with general training paradigms have been a) poor definitions of the training mechanism, b) poor design of training protocols as well as c) being relatively underpowered to detect small effects (Smid et al., 2020). In the current study, these issues are addressed in the training paradigm in several ways; 1) it employed a focused training paradigm that targeted the mechanism of interest, inhibition, 2) the training employed an adaptive and double-blind randomized control trial, 3) The current study was conducted on a relatively large developmental sample.

In addition, non-significant results in training studies can mean that there is evidence for the null hypothesis (training does not affect the measure of interest) or that the data are insensitive to providing evidence supporting the null hypothesis (Dienes, 2014; Dougherty et al., 2016). It is, therefore, of interest in training studies to consider evidence in favor of the null hypothesis in particular. To this end, the current study uses the Bayes factor in support of the null hypothesis (cognitive control training did not lead to changes in the decision-making measures) to assess training-related transfer to decision-making tasks. In this case, the Bayes Factor *B* indicates that the data are *B* times more likely under the null model than the training model. Therefore, the Bayes Factor allows three different types of conclusions regarding the potential training effects; (i) strong evidence for a null effect (no-training related changes) (*B* much greater than 1), (ii) strong evidence for the training model (*B* close to 0); (iii) and the evidence is insensitive (*B* close to 1) (Dienes, 2014).

Here, in a large sample of children aged 6 to 12 years (N = 205), I research the relationship between latent factors of cognitive control and decision-making. In addition, I investigate both short- and long-term effects of training EFs in the form of inhibitory control on pro-social and intertemporal decision-making. The current training paradigm follows the gold-standard approach for a well-powered, rigorous, and double-blind, randomized controlled trial (RCT) (Smid et al., 2020). Potentially, younger children may benefit more from the training paradigm (Wass et al., 2012), which is why I include age-related training changes. Nevertheless, the results suggest that cognitive control training targeting inhibition does not change pro-social or intertemporal decision-making in middle childhood.

4.3 Methods

4.3.1 Participants

A total of 262 typically developing children were recruited for the study from schools within Greater London in the United Kingdom (data collection started in May 2019 and ended in May 2021). A final sample of 205 participants (mean age = 8.97, age range 6.03 – 12.61 years, 53.2% girls) was available who had (a) at least completed one decision-making task at any time point (N = 229) and

(b) also completed at least one successful valid training session (N = 205). Training sessions were valid when participants successfully completed at least two of the seven types of training games in a session. This led to the exclusion of 24 participants who did not complete a valid training session.

In-person (offline) testing took place at UCL campus testing facilities, and participants completed these tasks as part of a larger test battery. The decision-making games included a Dictator Game, an Ultimatum Game, and an intertemporal choice and valuation task. For offline testing, participants were guided through the battery by an experimenter. Participants completed the tasks on a laptop (Lenovo, running Windows 10), which were programmed in MATLAB or EPrime. Due to Covid, testing was moved online, which was conducted on the Pavlovia platform, and the tasks were coded in Python and JavaScript. Participants completed these at home.

For online testing, the participants' parents were requested to guide them through the task's instructions but not interfere with their behavioral performance on the task. The online battery consisted of detailed instructions recorded via audio and accompanied by slides with information, which participants had to listen to in full before performing the tasks. Further, comprehension questions ensured that participants had fully understood the tasks before taking part. Testing medium (Online / Offline) was included in the statistical models to control for potential effects.

4.3.2 Cognitive control training paradigm

Participants were randomly assigned to an Experimental Group (training inhibition; *N* = 107) or a Control Group (training response speed; *N* = 98). The training program consisted of an eight-week intervention where participants completed four training sessions per week, each lasting 15 minutes. Within each training week, one session took place at the children's school and was supervised by the experimenters, whereas the remaining three sessions took place at home. Parents of the participants were encouraged to oversee the training (note that, for children who enrolled in the study after the outbreak of the COVID-19 pandemic in March 2020, all training sessions took place at home). The training was computerized and used a highly motivating and gamified interface and narrative. Participants were told that they were a pilot that crashed on an island and that they had to navigate through different environments to collect parts of their broken airplane, after which they could return home. For more details on the training, see 4.6.1 Cognitive control training protocol.

		6	7	8	9	10	11	12+
Control	Offline	5	14	14	21	21	9	3
	Online	0	0	2	2	3	3	0
Experimental	Offline	9	14	14	19	24	11	3
	Online	0	0	3	4	1	2	3

Figure 39. Participants in Training Paradigm.
Participant numbers across training groups (control and experimental groups), testing medium (offline testing in purple and online testing in blue) and across years of age.

Moreover, the training was adaptive to each child's performance to avoid ceiling and floor effects and keep children motivated throughout the sessions. Seven training games were randomly assigned to the sessions so that participants would play a different set of games in each session (around three games per session). The games took place in different settings (e.g., forests, deserts, tundra, or mountains) and required participants to gain points by collecting treasures, gems, or coins while avoiding a perpetrator (e.g., dragon, monster, or ghost) (see Supplementary Figure 50).

While the training games were presented in the same manner across both groups, i.e., the stimuli and appearance of the games were the same, the instructions given to each group varied according to the abilities being trained. The experimental group underwent response inhibition training, where the stop-signal was implemented in the context of the training games. Different stimuli were used as go- and stop-signals depending on the game (Figure 40d). Participants in the experimental group were instructed to react as quickly as possible if the go-signal appeared. However, they were instructed to withhold a response if the stop-signal appeared (26-47% of total trials depending on the game, mean = 32%), thus requiring them to inhibit the go-signal response. The stop-signal delay (SSD; i.e., the delay between the presentation of the go-signal and the stop-signal) was initially set to 200ms and was adjusted to participants' performance using an adaptive staircase procedure. If participants successfully inhibited their response, then the SSD was increased by 50ms to make the task more difficult; however, if participants did not inhibit their response, then the SSD was decreased by 50ms to make the task easier. This ensured that the training was adaptive and avoided floor or ceiling effects.

The response speed training was identical to the experimental condition in all aspects except that a response was required for all signals. Participants were instructed to press the spacebar as quickly as possible. To ensure that training was adaptive for this group, participants had to respond within a time window that was set based on a rolling average of the response time of the previous ten trials plus two standard deviations. This ensured that the training was adaptive while minimizing the effect of outliers on the response threshold.

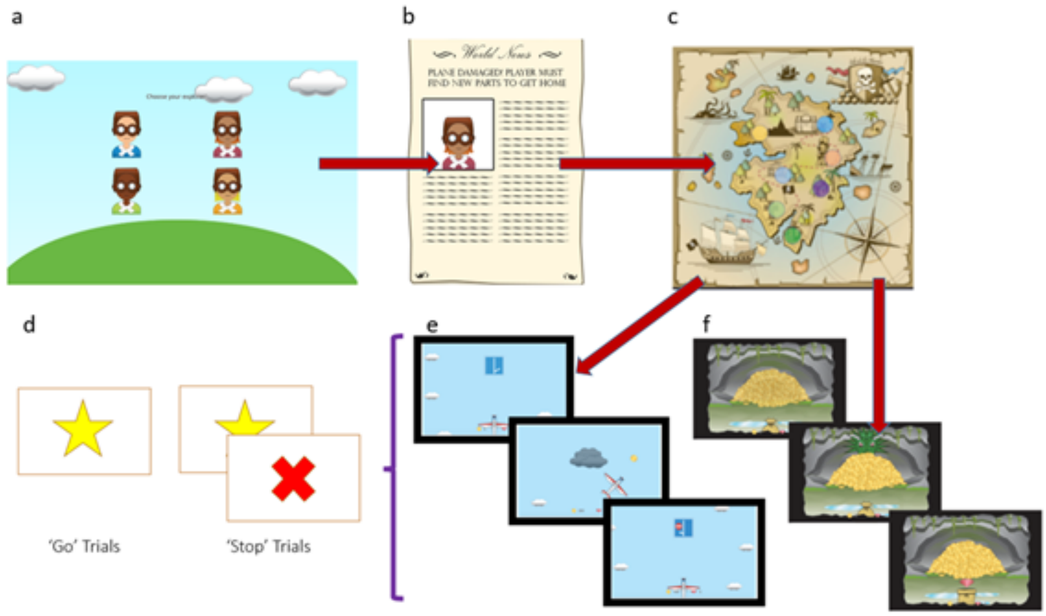


Figure 40. Cognitive control training paradigm.
Both groups were administered the same protocol, comprising a variety of adaptive and gamified tasks. (a-c) Children were told they were pilots who had crashed their plane on an island and had to navigate the island to earn coins. (d) Coins could be earned via seven unique games that operated on the same underlying mechanism: ' go' vs. 'stop trials. (e-f) Two examples of these games. Adapted from Ganesan et al. 2022.

4.3.3 Social and intertemporal decision-making tasks

Participants completed a cognitive battery with multiple tasks at each time point. The first stage of the battery consisted of eight EF tasks (Table 6), after which participants completed the decision-making tasks as the final part of the battery. For offline testing, participants were guided through the testing battery by an experimenter and completed comprehension checks. For the Dictator Game, participants were walked through three different scenarios of distributing the

MUs; keeping all 6 MUs, giving away all 6 MUs, or splitting the MUs evenly. They were prompted to answer the consequences of each scenario (e.g. “If you chose to keep all 6 coins, how many do you have? How many does the other child have?”). Similar comprehension checks were used for the Ultimatum Game and the intertemporal choice tasks (e.g. “If you choose the option with 6 moons and 6 coins, how long do you have to wait until you get your present?”). If they responded incorrectly, they were corrected by the experimenter. For online testing, the same comprehension checks were conducted, and the children were prompted for their answer, but here they were given the answer, after which the answer was briefly explained (e.g. “If you choose to keep 3 coins and give 3 coins away, you and the other child will both have 3 coins), but not checked for comprehension further as there was no experimenter present.

4.3.3.1 Dictator Game

Participants were allocated six monetary units (MUs), visually represented in the task as coins, which could be exchanged for gifts at the end of the experiment. Participants were told that their collected MUs would go towards their present at the end of the testing session and that the more MUs they had at the end of all the games, the larger their gift would be. The reward was described in this abstract way to appeal to children of all ages and has previously been found to be incentivizing for children of this age range (Smid et al., 2022; Steinbeis et al., 2016). In the offline sample, two boxes were presented, one for the child and one for their “partner.” Children were told that they were playing with another child from a different school; in reality, there was no other participant. They were instructed to first click on the MU and then the boxes to divide them (Figure 41a); they were also informed that once they had put a MU in a box, they could not change their decision. Counters at the side of the boxes kept track of the number of MUs in either box. During the task, the instructor explicitly informed the participant that they would turn away and not look at the screen so that participants were free to decide on distributions without judgment. There was no response time limit for the participants. The Dictator Game (DG) thus measures pro-social decision-making as indicated by how many MUs a participant decided to give away to another unknown child.

In the online version, children determined their chosen distribution by moving a slider (Figure 41b). In this sense, the online task required just one move to distribute the MUs. As in the offline version, children were told that they were playing with another child from another school whom they did not know when in reality, there was no other participant. Unlike in the offline sample, however, children could change their minds about their preferred distributions indefinitely and submit their final decision by pressing the spacebars on their computers. Parents were instructed to be present in the room while testing, engaged in an activity such as reading a book, and not to influence their children's participation.

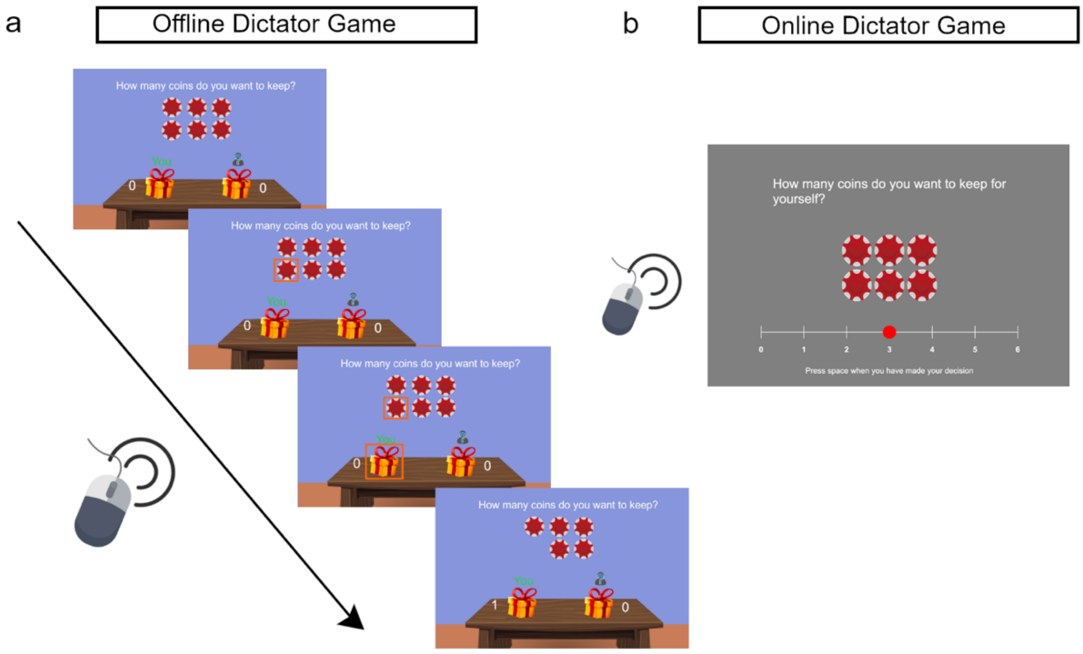


Figure 41. Dictator Game (DG) across the offline and online testing mediums.

(a) the offline version of the DG required participants to click on one MU (coin) at a time and then on the desired box (depicted as a present), which they wanted to deposit the MUs in. Participants had to do this for every MU and were given as much time as needed while the experimenter turned around and looked away. (b) the online version of the DG differed from the offline version in that participants had to click on a slider system to determine how many MUs they wanted to keep for themselves. They were allowed as much time as possible and could click different options on the slider as often as they wanted. As testing took place at home, parents were instructed to look away while the participants completed this task.

4.3.3.2 Ultimatum Game

The Ultimatum Game (UG) consisted of two parts, a proposer and a responder role. Like in the DG, in the UG proposer role, children had to distribute six MUs amongst themselves and another unknown child taking part in the study. However, this time, they were told that the other child had the option to reject their offer. If the other child rejected their offer, the participant and the other unknown child would both receive zero MUs. There was again no limit to their response time. The second part of the UG was the UG responder role, where children could accept or reject a single offer of an unfair distribution (1/6) of MUs made by another unknown child in the study (Figure 42a, b). If they rejected the offer, the participant and the unknown other child who made the offer (a computer in reality) would receive zero MUs. For this game, there was again no response limit for the participants.

Due to time restrictions, the UG Proposer role was only conducted offline and not included in the online version of the battery. Therefore, data from the 36 online participants was missing. These missing data were imputed; for details on the imputation procedure, see 4.6.2 Imputation of missing data.

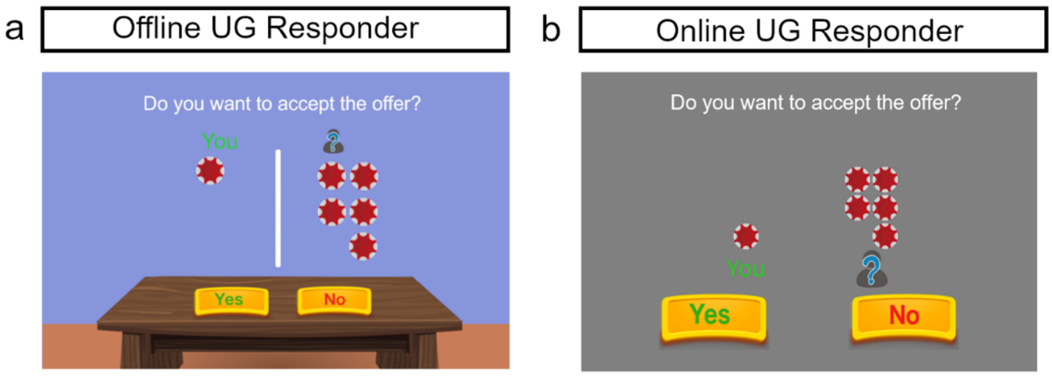


Figure 42. Ultimatum Game (UG) Responder, across offline and online testing mediums.

(a) for the offline UG, participants had to click either a button with 'Yes' to accept the offer or 'No' to reject the offer. (b) The online version of the UG Responder role was nearly identical in that online participants also had to click a button with 'Yes' or 'No' to accept or reject the offer. There was no limit to a participant's response time.

4.3.3.3 Intertemporal choice task

Intertemporal decision-making was assessed using two tasks, an intertemporal choice and an intertemporal valuation task. In the intertemporal choice task, participants made choices between immediate and delayed reward options. This task measured the extent to which participants would discount via choices between the two options. Participants completed 18 trials (in a fixed order) where they were always presented with a choice between either an immediate or a delayed option (Figure 43a, b). The value of delay used was days, where every moon depicted indicated one additional day of waiting before the participant would receive their reward. The reward for the delayed option was always eight MUs, and the immediate reward option ranged from two MUs to four MUs and then six MUs. Participants' discounting was measured by calculating the percentage of total delayed choices.

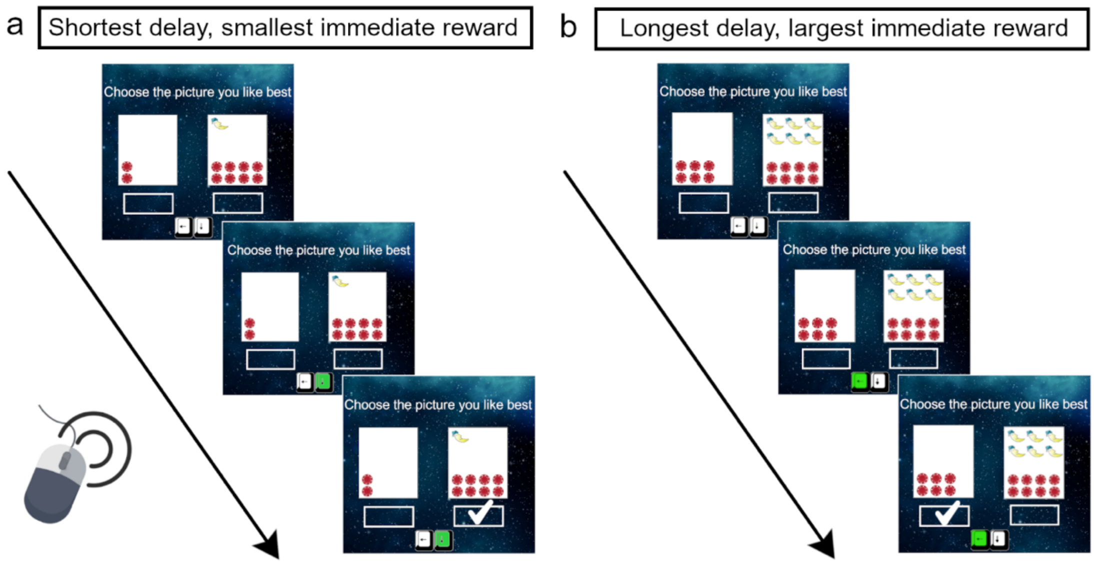


Figure 43. Intertemporal choice trials.
Participants were presented with two choices and had to use the left and down arrow keys to decide. (a) the first trial of the task, where the immediate reward was small (two MUs), and the largest reward was presented for the shortest delay (one day). (b) the last trial of the task, where participants had to choose between the largest immediate reward (six MUs) and the large reward for the longest delay (six days).

4.3.3.4 Intertemporal valuation task

Participants rated all the different delay and immediate reward options for the intertemporal valuation task, which continuously measured the explicit valuation of delay and reward. Participants completed six trials, one for each delay option ranging from one unit of delay to six units of delay. The reward was always eight MUs, which was the highest available reward option. Ratings were indicated on a continuous scale which ranged from -5 to +5 (Figure 44a). Afterward, participants explicatively rated how much they valued immediate reward, ranging from two MUs to four MUs, six MUs, or eight MUs (Figure 44b). For this task, a linear slope was fitted to each participant's continuous ratings for the delay and reward trials separately to measure their explicit delay discounting and reward valuation. As individual measures of delay discounting, the steepness of the linear slopes was extracted for each participant (the coefficients of the linear slope), as well as the intercept, which reflects the initial valuation of eight MUs for the smallest amount of delay (one day). For the reward valuation trials, the steepness of the slope for the valuations (the coefficients) was extracted, as well as the initial valuation of the smallest immediate offer (two MUs immediately).

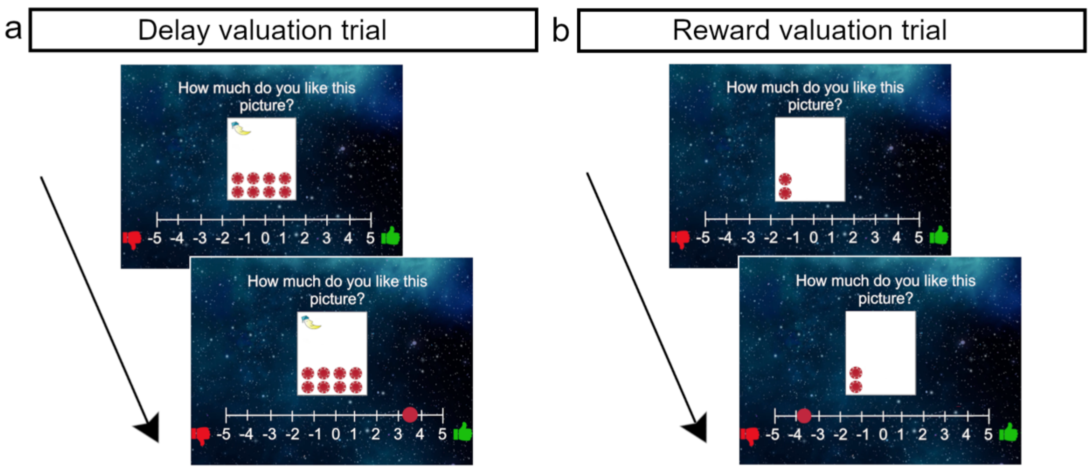


Figure 44. Delay devaluation and reward valuation trials.

(a) valuation trials for delay, where participants rated how much they valued the large immediate reward (eight MUs) for increasing delays (one day to six days). (b) valuation trials for reward, where participants rated how much they valued increasing values of immediate reward (two MUs to eight MUs).

The coefficients and intercepts were combined in a weighted average, which reflects each participant's continuous rating of delay and reward. The weighted averages for delay and reward were significantly correlated to the separate intercept and coefficients, indicating that these measures managed to capture both independent measures. For more details on how these measures were derived, see 4.6.4 Linear valuation composite measures.

4.3.4 Cognitive control latent factors

As participants completed many EF measures, I used latent factors of EFs to reduce measurement related variance as well as the number of tests to run. To obtain latent factors of EFs, a confirmatory factor analysis (CFA) was conducted using Lavaan in R on eight measures from a broad EF battery (see 3.3.3 Cognitive task battery and Table 6) (Rosseel, 2012). This provided three factors that loaded on Inhibition ("I"), Cognitive Flexibility ("S"), and Working Memory ("M"). These factors were coded so that when the value of a factor was higher, this reflected a better score on the respective EF. For more details on the factor structure, see 4.6.3 Confirmatory Factor Analysis for executive measures.

4.3.5 Statistical methods and imputations

Missing data were imputed using the MICE package in R (50 datasets created, 50 maximum iterations), and Quickpred was used to generate the imputation model. For details on the missing data and the imputation performance, see 4.6.2 Imputation of missing data. For all main results in the current chapter, I report the results from the pooled models over all imputed datasets. Linear mixed models were used to assess training effects with the lme4 package (Bates et al., 2015). For the plots involving age and the decision-making measures, the effect of the testing medium was regressed out of these relationships, and the residuals were plotted. Reaction times were log-transformed before analysis, and extreme outliers were removed from the reaction time data, leading to the exclusion of one data point for the average reaction time for the intertemporal choice task.

For evidence of null effects, I report the Bayes factor in favor of the null model (the model without a group-by-session interaction) over the training model (model with a group-by-session interaction) for each measure of interest (Dienes, 2014). I isolate this particular interaction as the training effect of interest. In this case, the Bayes Factor *B* indicates that the data are *B* times more likely under the null model than the training model. Therefore, the Bayes Factor allows three different types of conclusions regarding the potential training effects; (i) strong evidence for a null effect (no-training related changes) (*B* much greater than 1), (ii) strong evidence for the training model (*B* close to 0); (iii) and the evidence is insensitive (*B* close to 1) (Dienes, 2014). To help interpretation, I use conventional cut-offs for the Bayes Factor; a *B* greater than 3 represents substantial evidence for the null model, a *B* of 1/3 (0.33) provides substantial evidence for the training model, and a *B* between 3 and 1/3 provides support for neither and may indicate data insensitivity (Jeffreys, 1961). Bayesian models were run using the BRMS package in R and in parallel sessions using the Future package (Bengtsson, 2021; Bürkner, 2017). Outcome variables were standardized, and models were run with lazy normal priors (mean of 0 and standard deviation of 1), with adapt delta set between 0.8 and 0.9 and the number of iterations between 10,000 and 20,000 depending on convergence.

The data and code used to conduct the analyses reported in this chapter can be found on my Github: https://github.com/ClaireSmid/CogControlTraining_DecisionMaking

4.4 Results

4.4.1 Developmental effects in social and intertemporal decision-making

I corrected the p-values of the age-related effects for multiple comparisons using a Bonferroni correction, with the threshold for significance set to p = .006. For the mean values and standard deviations, see Table 7.

Table 7. Social and intertemporal decision-making mean values at pre-training.

Task Name	Mean Values	Standard Deviation
Dictator Game (MUs given (max 6))	2.12	1.07
Ultimatum Game (Proposer) (MU's given (max 6))	2.78	0.70
Ultimatum Game (Responder) (probability to accept)	0.41	0.49
Strategic Decision Making (difference in MUs given)	0.65	1.03
Percentage of delayed choices	37.95%	32.33%
Delay valuation (weighted average)	0.67	1.44
Reward valuation (weighted average)	-0.51	1.58

4.4.1.1 Social decision-making

For the Dictator Game (DG), the amount of MUs given was positively correlated to age (β = 0.15, se = 0.04, t = 3.41, p = .001), and older children had slower reaction times for distributing the DG MUs (β = 0.14, se = 0.02, t = 6.28, p < .001). In addition, strategic decision-making (the difference between MUs given in the DG and UG) was significantly negatively related to age after correction (β = -0.16, se = 0.05, t = -3.46, p = .001). Thus, older children would initially offer more MUs during the dictator game and were slower to distribute their MUs during the Dictator Game. Their offers stayed more consistent through to the Ultimatum Game.

There was no correlation with age for the probability to accept the UG Responder unfair offer (β = -0.03, se = 0.02, t = -1.18, p = .240), however, the decision time for responding to the offer was positively correlated to age (β = 0.12, se = 0.03, t = 3.66, p < .001). Thus, reaction times increased with age for both distributing the MUs and deciding how to respond to the unfair offer.

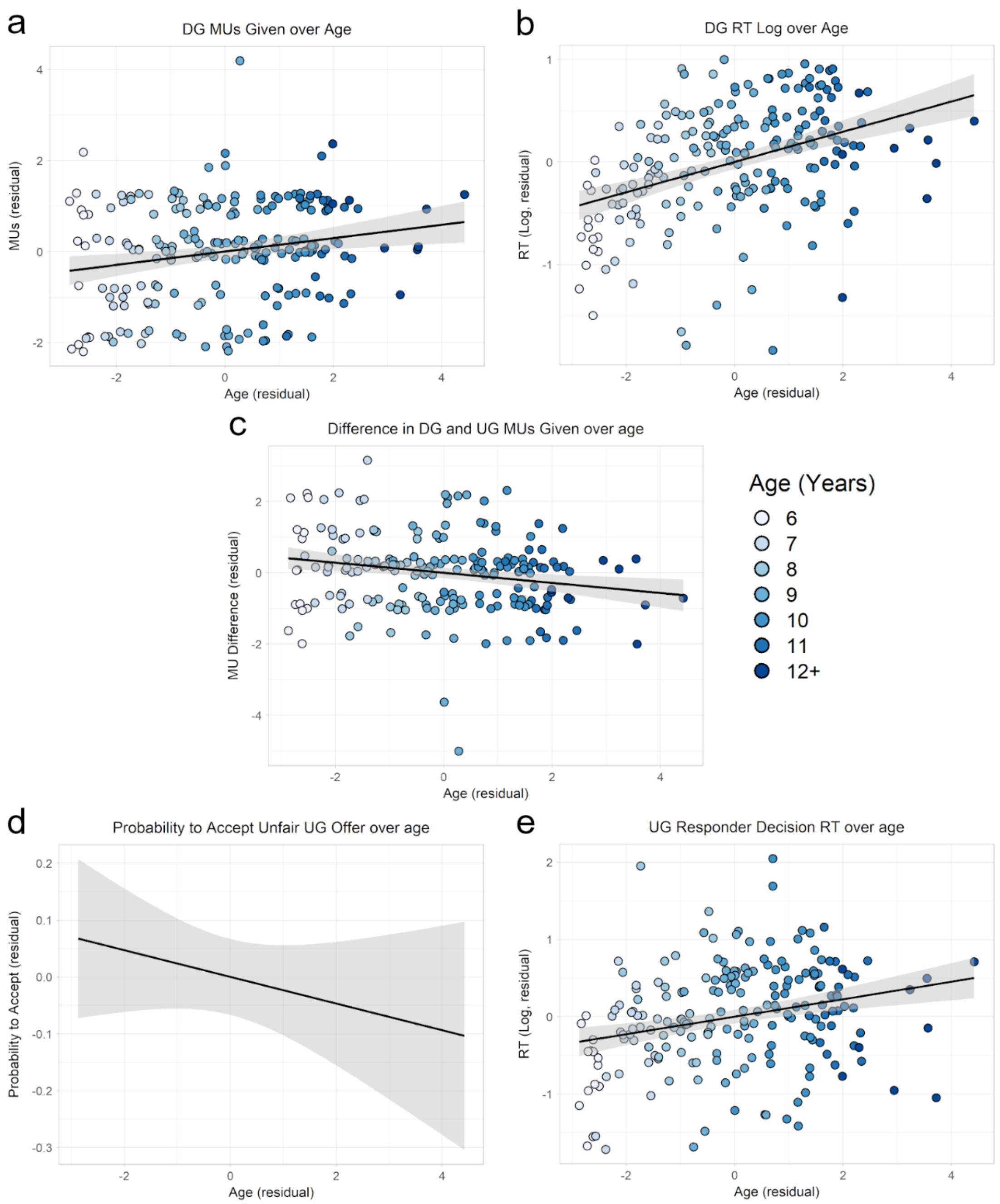


Figure 45. Social decision-making measures over age, controlled by testing medium.

(a) the number of MUs given during the DG increased with age, (b) and so did the reaction time for the DG. (c) strategic decision-making (the difference between offers made in the DG and UG Proposer) decreased with age. (e) the probability of accepting the unfair offer in the UG Responder was not significantly correlated with age, (f) while the reaction time for this decision again increased with age. Note: These plots were based on the first imputed dataset only.

4.4.1.2 Intertemporal decision-making

There was no significant relationship between age and the total percentage of delayed choices (β = 0.02, se = 0.04, t = 0.45, p = .652), and no significant correlation between age and the average reaction time (β = 0.03, se = 0.03, t = 0.85, p = .396). One outlier in the average reaction time for outliers was found and removed. When leaving the outlier in, there was still no significant correlation between age and the average reaction time (β = 0.03, se = 0.03, t = 0.89, p = .377).

For delay devaluation, there was a significant positive relationship with age (β = 0.23, se = 0.05, t = 4.39, p < .001), indicating that with age, children valued increases in delay as progressively worse. For reward valuation, there was no significant relationship with age (β = 0.08, se = 0.04, t = 1.86, p = .065), indicating that children did not value increasing reward more with age.

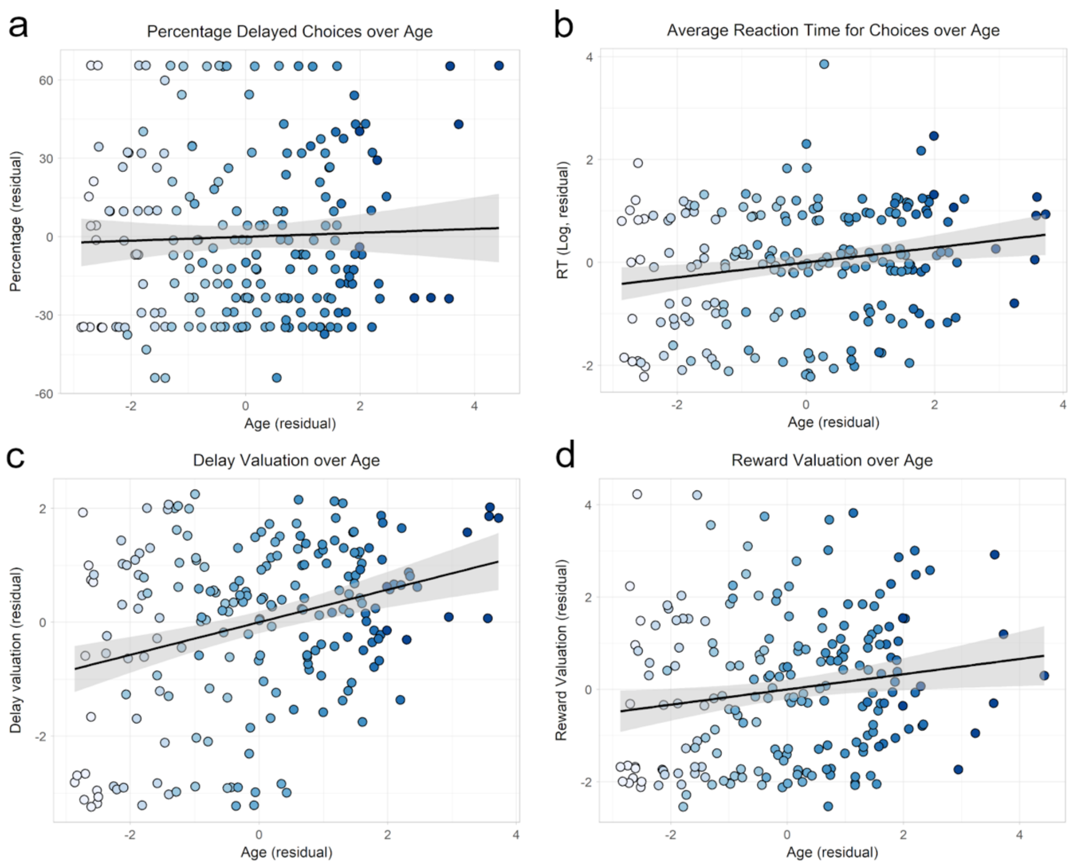


Figure 46. Intertemporal decision-making measures and age controlled by testing medium.

(a) the percentage of delayed choices did not change over age (b), and neither did the average reaction time for choices (c); however, delay increased with age (d) while there was no significant change in reward valuation over age. Note: these plots were based on the first imputed dataset; one outlier was removed from the average reaction times for the intertemporal choice task.

4.4.2 Relationships between executive functions and decision-making measures

First, I investigated whether there were correlations between social and intertemporal decision-making and the latent factors of cognitive control at the pre-training time point. Next, I corrected these potential relationships for age.

4.4.2.1 Inhibition factor

The latent inhibition factor was significantly positively correlated to the amount of MUs given during the Dictator Game (β = 0.65, se = 0.26, t = 2.44, p = .015), and negatively correlated to strategic decision-making (β = -0.81, se = 0.30, t = -2.71, p = .007). When correcting for age, these relationships were no longer significant (Dictator Game Coins: β = 0.34, se = 0.29, t = 1.19, p = .235; Strategic Decision-making: β = -0.45, se = 0.32, t = -1.41, p = .162). There were no significant relationships between inhibition and the responder behavior during the Ultimatum Game (β = -0.08, se = 0.14, t = -0.59, p = .554).

For intertemporal decision-making, there was no correlation with delayed choices (β = 0.10, se = 0.28, t = 0.37, p = .710). There was a significant correlation between the latent inhibition factor and delay valuation (β = 1.10, se = 0.34, t = 3.25, p = .001), and no significant relationship with reward valuation (β = -0.01, se = 0.28, t = -0.04, p = .970). When correcting for age, the relationship between the latent inhibition factor and delay valuation was no longer significant (β = 0.51, se = 0.36, t = 1.43, p = .156).

4.4.2.2 Cognitive flexibility factor

There was no significant correlation between the latent cognitive flexibility factor and the amount of MUs given during the Dictator Game (β = 1.84, se = 1.24, t = 1.49, p = .139), but it was significantly negatively correlated to strategic decision-making (β = -3.39, se = 1.44, t = -2.36, p = .020), suggesting that worse cognitive flexibility was related to more strategic decision-making. When correcting for age, the relationship between strategic decision-making and the latent cognitive flexibility factor was no longer significant (β = -1.27, se = 1.67, t = -0.76, p = .448). There was no significant relationship with the probability to accept the offer during the Ultimatum Game Responder (β = 0.21, se = 0.62, t = 0.34, p = .733).

For intertemporal decision-making, there was a significant positive relationship between the latent cognitive flexibility factor and the percentage of delayed choices (β = 3.02, se = 1.24, t = 2.43, p = .016) as well as delay valuation (β = 5.13, se = 1.21, t = 4.23, p < .001), indicating that more cognitive flexibility was related to a higher percentage of delayed choices and stronger delay valuation. When correcting for age, this relationship remained significant for both measures (Percentage of delayed choices: β = 3.33, se = 1.45, t = 2.30, p = .022; Delay valuation: β = 2.81, se = 1.38, t = 2.04, p = .043). There was no significant relationship between reward valuation and the factor (β = 1.75, se = 1.23, t = 1.42, p = .156).

4.4.2.3 Working memory factor

The latent working memory factor was significantly positively correlated to the amount of MUs given during the Dictator Game (β = 4.86, se = 2.26, t = 2.15, p = .033), and negatively to strategic decision-making (β = -7.20, se = 2.52, t = -2.86, p = .005). Suggesting that better working memory was related to a higher number of MUs given during the dictator game and to less strategic decision-making. When correcting for age, the relationship between MUs given during the Dictator Game and the working memory factor was no longer significant (β = 1.41, se = 2.63, t = 0.54, p = .592), or strategic decision-making (β = -3.61, se = 2.91, t = -1.24, p = .216). There was no significant relationship between working memory and the probability of accepting the offer during the Ultimatum Game Responder (β = -0.53, se = 1.14, t = -0.47, p = .642).

For intertemporal decision-making, there was no significant relationship between working memory and the percentage of delayed choices (β = 3.05, se = 2.32, t = 1.32, p = .189), but there was a significant relationship with delay valuation (β = 8.99, se = 2.26, t = 3.98, $p < .001$). When correcting for age, the relationship between delay valuation and the working memory factor was no longer significant (β = 4.51, se = 2.56, t = 1.76, p = .080). There was no significant relationship between working memory and reward valuation (β = 1.77, se = 2.28, t = 0.77, p = .440).

We also assessed inter correlations between the decision-making task measures and whether social and intertemporal decision-making could be captured by a single general factor, similar to a previous study with young adults that identified a general factor for decision-making (Moutoussis et al., 2021). I did not find that a single decision-making factor explained behavior across the two decision-making paradigms, and I, therefore, assessed the separate measures of the decision-making measures. For the factor analysis and the correlation between the decision-making measures, see 4.6.5 Decision-making factor analysis.

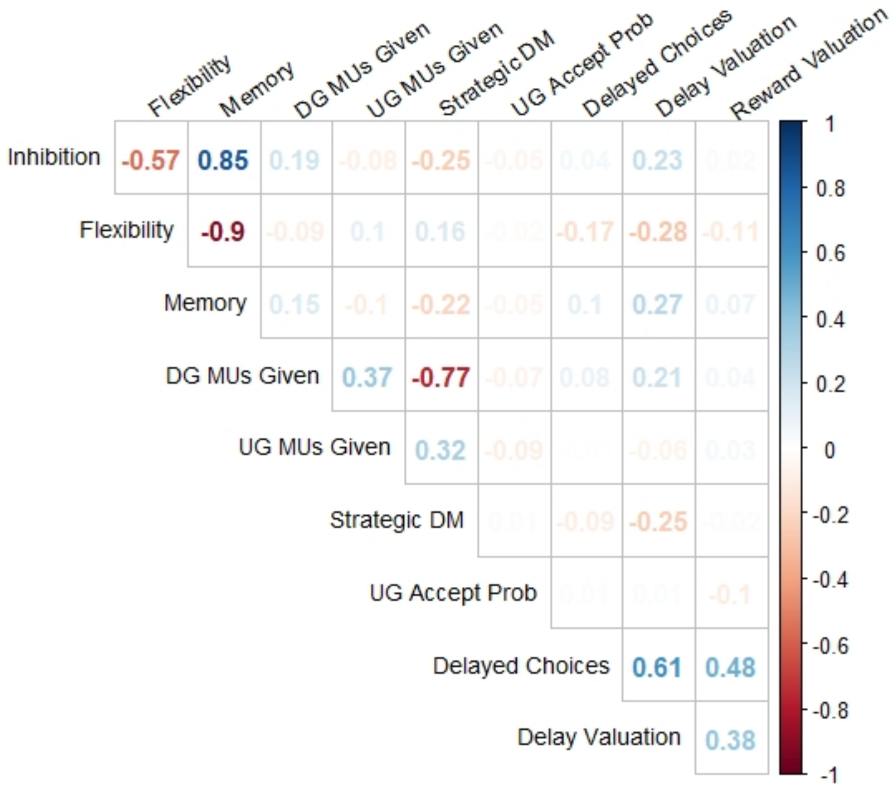


Figure 47. Correlations between EF factors and decision-making measures.

4.4.3 Short-term training effects

Next, I investigated short-term training effects between the pre-training and post-training time points immediately after training completion. Potential training effects were assessed with linear mixed models (e.g., (y ~ group * session + testing medium + (1|ID)). I focus on the essential decision-making measures, which for social decision-making reflect pro-social decision-making (Dictator Game MUs given), strategic social decision-making (difference between Dictator Game and Ultimatum Proposer Game MUs given), inequality aversion (probability to accept the unfair offer in the Ultimatum Responder Game). For intertemporal decision-making, I focus on intertemporal choice (percentage of delayed choices in the intertemporal choice game), delay valuation (how strongly participants devalued the same reward for increasing delay), and reward valuation (how strongly participants valued increases in reward). The Bayes Factors for all measures are reported in Table 8. I also report the Bayes Factors for the null model (*BF_{01},* no interaction between session and group) in the text for each measure.

4.4.3.1 Social decision-making

For the role of the proposer in the Dictator Game, there was no significant main effect of session (beta = 0.03, se = 0.14, t = 0.21, p = .831), nor of group (beta = 0.26, se = 0.14, t = 1.86, p = .064). There was also no significant group by session interaction (beta = 0.01, se = 0.18, t = 0.08, p = .940, BF_{01} = 5.44; see Figure 6a).

For strategic social decision-making (i.e., difference in offer between the Ultimatum and Dictator Games), there was no significant main effect of session (beta = -0.09, se = 0.17, t = -0.50, p = .615), while there was a main significant effect of group (beta = -0.34, se = 0.16, t = -2.16, p = .031). There was no group by session interaction (beta = 0.04, se = 0.22, t = 0.30, p = .841, BF_{01} = 1.28, see Figure 6b).

For the role of the responder in the Ultimatum Game, there were no main effects of session (beta = -0.40, se = 0.39, z = -1.03, p = .305), or group (beta = -0.13, se = 0.38, z = -0.34, p = .733, see Figure 6c). There was also no significant interaction between session and group (beta = 0.90, se = 0.54, z = 1.68, p = .094, BF_{01} = 6.03).

4.4.3.2 Intertemporal decision-making

For the total percentage of delayed choices, there was no main effect of session (beta = 0.06, se = 0.14, t = 0.41, p = .677) or group (beta = 0.11, se = 0.14, t =

0.76, p = .448), or a session by group interaction (beta = 0.02, se = 0.19, t = 0.12, p = .901, BF_{01} = 4.17, see Figure 6d).

For delay valuation, there was a significant effect of session (beta = 0.38, se = 0.18, t = 2.16, p = .032), but no significant effect of group (beta = 0.02, se = 0.16, t = 0.15, p = .880), or a session by group interaction (beta = -0.11, se = 0.24, t = -0.45, p = .654, BF_{01} = 3.59, see Figure 6e).

For reward valuation, there was no significant effect of session (beta = 0.34, se = 0.18, t = 1.85, p = .066), nor of group (beta = 0.03, se = 0.16, t = 0.22, p = .829) nor a session by group interaction (beta = -0.13, se = 0.25, t = -0.53, p = .596, BF_{01} = 1.49, see Figure 6f).

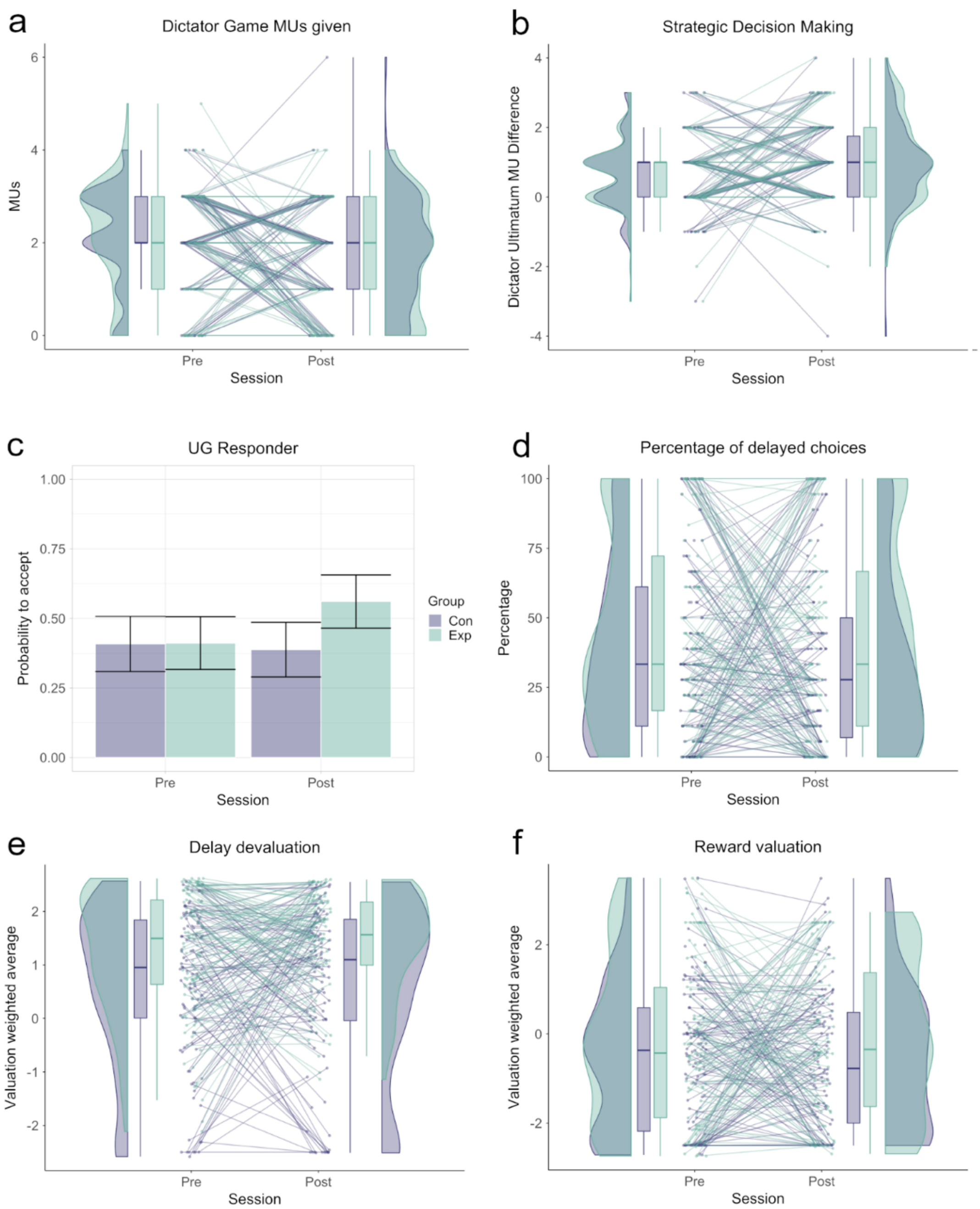


Figure 48. Pre- and post-training related changes in social and intertemporal decision-making for experimental and control groups.

Social decision-making measures include (a) the number of MUs given during the Dictator game, (b) strategic decision-making, which is the difference in coins given for the Dictator and Ultimatum Proposer Game, and (c) the probability of accepting the unfair offer in the Ultimatum Responder Game. Intertemporal decision-making measures include (d) the total percentage of delayed choices during the intertemporal choice game, (e) the discounting of delay, and (f) the valuation of reward. Note: to visualize the data, a single imputed dataset was used.

4.4.4 Long-term training effects

I also investigated whether any long-term effects emerged as a function of training by comparing performance before training to performance one-year after the end of training (i.e., follow-up).

4.4.4.1 Social decision-making

For the role of the proposer in the Dictator Game, there was no significant main effect of session (beta = 0.18, se = 0.12, t = 1.43, p = .154), nor of group (beta = 0.26, se = 0.14, t = 1.84, p = .067). There was also no significant group by session interaction (beta = -0.15, se = 0.10, t = -1.48, p = .139, BF_{01} = 5.21, see Figure 7a).

For strategic social decision-making, there was no significant main effect of session (beta = -0.17, se = 0.21, t = -0.80, p = .426), but there was a significant main effect of group (beta = -0.34, se = 0.16, t = -2.14, p = .033). There was no significant group by session interaction (beta = 0.18, se = 0.14, t = 1.32, p = .190, BF_{01} = 0.75, see Figure 7b).

For the role of the responder in the Ultimatum Game, there was no significant main effect of session (beta = -0.35, se = 0.29, z = -1.24, p = .217), or group (beta = -0.12, se = 0.34, z = -0.35, p = .730). There was also no significant group by session interaction (beta = 0.22, se = 0.24, z = 0.91, p = .365, BF_{01} = 33.31, see Figure 7c).

4.4.4.2 Intertemporal decision-making

For the total percentage of delayed choices, there was no main effect of session (beta = 0.08, se = 0.12, t = 0.72, p = .474) or group (beta = 0.11, se = 0.14, t =

0.77, p = .442), nor a session by group interaction (beta = -0.39, se = 0.09, t = -0.42, p = .677, BF_{01} = 7.79, see Figure 7d).

For delay valuation, there was no significant effect of session (beta = -0.02, se = 0.12, t = 0.19, p = .851) or group (beta = 0.02, se = 0.14, t = 0.13, p = .895), nor a session by group interaction (beta = -0.01, se = 0.10, t = -0.11, p = .910, BF_{01} = 14.94, see Figure 7e).

For reward valuation, there was no significant effect of session (beta = -0.01, se = 0.13, t = -0.07, p = .944), or group (beta = 0.03, se = 0.16, t = 0.20, p = .838), nor a session by group interaction (beta = -0.06, se = 0.11, t = -0.59, p = .554, BF_{01} = 6.84, see Figure 7f).

Table 8 displays the Bayes Factors for the short- and long-term training-related effects.

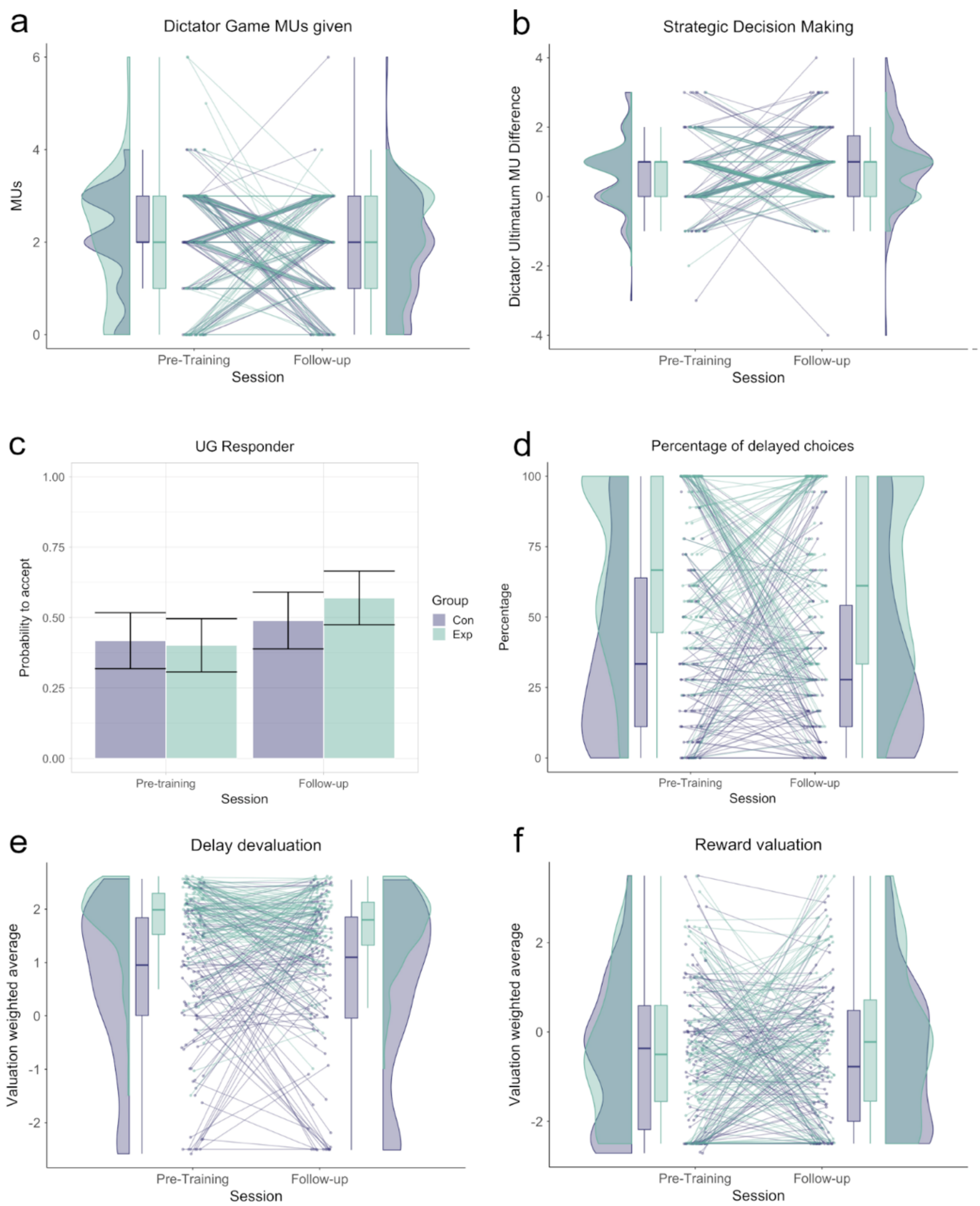


Figure 49. Pre-training and one-year follow-up related changes in social and intertemporal decision-making for the experimental and control groups.

Social decision-making measures include (a) the number of MUs given during the Dictator game, (b) strategic decision-making, which is the difference in coins given for the Dictator and Ultimatum Proposer Game, and (c) the probability of accepting the unfair offer in the Ultimatum Responder Game. Intertemporal decision-making measures include (d) the total percentage of delayed choices during the intertemporal choice game, (e) the discounting of delay, and (f) the valuation of reward. Note: to visualize the data, a single imputed dataset was used.

Table 8. Bayes Factors for short-term and long-term training effects.

Bayes Factors were acquired by running the same model on all imputed datasets, and results were then combined in one fitted model object. Conventional cut-off methods were used for interpretation, where B > 3 indicates strong support for the null, B < 1/3 means support for the alternative (training model), and 1/3 > B < 3 shows support for neither, reflecting data insensitivity.

Short-term training-related changes

P	0.907	0.432	0.073	0.749	0.812	0.456
B, giving support for						
Null	5.44		6.03	4.17	3.59	
Neither		1.28				1.49
Training						

Long-term training-related changes

P	0.145	0.161	0.314	0.910	0.745	0.481
B, giving support for:						
Null	5.21		33.31	7.79	14.94	6.84
Neither		0.75				
Training						

4.4.5 Age-related training effects

Lastly, I assessed whether training-related effects might be influenced by age.

To assess this, I entered continuous age into the linear mixed models (e.g., y ~

group * session * age + testing medium + (1|ID)). There were no significant age-related effects (e.g., an interaction between group * age, or group * session * age effects) for any of the measures, either pre-post training or pre-training to the one-year follow-up.

4.5 Discussion

In the current chapter, I investigated the relationship between pro-social and intertemporal decision-making and cognitive control by looking at correlations between decision-making and latent factors of cognitive control, as well as studying changes in social and intertemporal decision-making in a large (N = 205) sample of children aged 6-13 in the context of an eight-week training study. First, I investigated age-related changes and relationships to executive functions (EFs) with the decision-making measures at the pre-training time point. I observed that both pro-social and intertemporal decision-making increased with age. While I found several significant relationships between EFs and decision-making measures at the pre-training time point, these were not as strong as previous studies have reported. The only relationships that survived correction were the positive relationship between cognitive flexibility, intertemporal choice, and delay devaluation. Second, I investigated training effects across three time points (pre-training, post-training, and one-year follow-up). When I investigated training-related group differences, I found no differences after cognitive control training, which targeted inhibition in social and intertemporal decision-making in childhood, immediately after training, or at the one-year follow-up. I consider Bayesian evidence in favor of the null model (no training results), the training model, and data insensitivity and find that most evidence points towards an absence of training effects. I conclude that pro-social and intertemporal decision-making in this current sample does not seem to be trainable via cognitive control training.

First, I assessed whether I saw the expected patterns of age-related changes in the social and intertemporal decision-making measures. For pro-social decision-making, measured via the offer made in the Dictator Game, I saw that children increased their offers with age, which suggests they became more pro-social, similar to previous findings in other developmental studies (Bauer et al., 2014; Chajes et al., 2022; Fehr et al., 2008; McAuliffe et al., 2017). However, I saw that strategic decision-making, or the difference in offers made during the Dictator Game and Ultimatum Game Proposer, decreased with age, which contrasted with previous work that saw increases in strategic decision-making with age (Steinbeis et al., 2012). Thus, I observed that the offers of older children remained more constant across the two proposer games. In comparison, the younger children gave lower offers during the Dictator Game and then increased their offers during the Ultimatum Game Proposer to a fairer distribution. This indicates that the younger children in the sample displayed strategic decision-making. Thus, when faced with the potential consequence of punishment (the other child rejecting their offer, upon which both would receive nothing), they could adjust their decision accordingly.

Surprisingly, I saw no age-related effects on the likelihood of accepting or rejecting the Ultimatum Game Responder unfair offer. Here, children were faced with the decision to accept an unfair offer (five MUs for the other child and one MU for themselves), where if they rejected the offer, both children would receive nothing (in reality, the other child did not exist). A rejection of the participant in the Ultimatum Game Responder has been interpreted as social punishment for norm-violating behavior (making an unfair offer, indicating poor cooperation) (Guth et al., 1982; Steinbeis, 2016). However, alternatively, accepting the unfair offer could be seen as a logical decision, as by accepting the offer, the participant still receives a reward (one MU) rather than zero. Thus, even if both theories are valid, neither of these seemed to become apparent with age, and instead, other non-age-related factors may be of influence.

I did not observe strong age-related effects for the intertemporal choice and valuation task. Most importantly, there was no increase in the percentage of delayed choices made with age, in contrast to previous studies investigating intertemporal choice in childhood (Green et al., 1999; Prencipe et al., 2011; Steinbeis et al., 2016). However, in the delay devaluation task, where participants rated how much they valued the same larger reward (eight MUs) for increasing amounts of delay (one to six days), I saw that with age, children progressively rated increases in delay for the same reward as less favorable. For the reward valuation, where participants rated how much they valued increasing immediate reward (from two MUs to eight MUs), there was no increase in valuation with age. Thus, while the older children valued increases in delay as less favorable, this was not apparent from their choices.

For many of the decision-making tasks, older children took longer to make their decisions, which could reflect increased processing time. In addition, longer reaction times for the Dictator Game resulted in more MUs given away, and in the intertemporal choice task, longer reaction times led less steep temporal discounting. Both these relationships held after controlling for age.

These results are described in 4.6.5.3 Inter-correlations for social and intertemporal decision-making. Thus, in the current sample, longer reaction times were linked to more pro-social decisions and less steep temporal discounting, which suggests that children had to deliberate longer over these types of decisions, and they were not spontaneous or intuitive (Rand et al., 2012; Zaki & Mitchell, 2013). In contrast, the time taken to deliberate the unfair offer in the UG Responder task was not correlated to the likelihood of accepting or rejecting the offer.

Next, I assessed the relationship between the decision-making measures at the pre-training time point and EFs via three factors that captured inhibition, cognitive flexibility, and working memory. Several initial correlations were significant; for example, better inhibition and working memory were related to more MUs given during the Dictator Game, less strategic decision-making, and stronger delay devaluation. However, after correcting for age, these relationships were no longer significant. The only relationships that remained statistically significant were between cognitive flexibility, the percentage of delayed choices made in the intertemporal choice task, and the strength of delay depreciation. Thus, better cognitive flexibility was related to a higher percentage of delayed choices and the steeper devaluation of increasing delay for the same amount. Previous developmental studies found that working memory (Wesley & Bickel, 2014; Zhao et al., 2022) and cognitive control, i.e., inhibition (Figner et al., 2010; Steinbeis et al., 2012, 2016), were linked to less steep temporal discounting However, I did not replicate those findings in the current chapter. It should be mentioned, however, that the decision-making tasks and EF measures differed between these previous and current studies.

The main aim of the current chapter was to investigate whether cognitive control training targeting inhibition would lead to short-term and long-term changes in social and intertemporal decision-making. I did not find significant relationships between inhibition and the decision-making measures at the pre-training time point. However, significant correlations do not necessarily indicate that increasing one ability will lead to increases in the other. Instead, if the underlying mechanisms target the same functions, improvements in this mechanism may also lead to changes in the other function (Ganesan & Steinbeis, 2022; Smid et al., 2020). To this end, I investigated both short-term and long-term changes following training. I compared the decision-making measures at the pre-training time point to the measures immediately after completing the 8-week training paradigm. There were no significant group-by-session interactions for any of the measures, either for the short-term or long-term models.

A focus of the current chapter was assessing potential null effects relating to the training in the context of Bayesian evidence. In this case, I used the Bayes Factor in support of the null model to assess the training's effectiveness (Dougherty et al., 2016). The Bayes Factor could provide support for three types of conclusions; (i) strong evidence for a null effect (no-training related changes) (*B* much greater than 1), (ii) strong evidence for the training model (*B* close to 0); (iii) and the evidence is insensitive (*B* close to 1) (Dienes, 2014). In short, a *B* greater than 3 represents substantial evidence for the null model, a *B* of 1/3 (0.33) provides substantial evidence for the training model, and a *B* between 3 and 1/3 provides support for neither and may indicate data insensitivity (Jeffreys, 1961). When I assess the Bayes Factors for the short-term training-related changes, I see that the evidence does not support either the null or the training model for two measures, strategic decision-making and reward valuation. Therefore, I can conclude that the data is insensitive to make comprehensive statements about these two measures. However, the Bayesian evidence points towards the null model as the most likely explanation for the other decision-making measures, e.g., pro-social (the Dictator Game MUs) and intertemporal decision-making (the percentage of delayed choices). Thus, I can conclude that from both the short- and long-term training-related changes, cognitive control training did not enhance pro-social and intertemporal decision-making as measured in this chapter.

The absence of any training-related effects is unexpected, as previous literature has suggested that EFs, especially relating to inhibitory control, may be underlying both social and intertemporal decision-making (Figner et al., 2010; Knoch et al., 2006; Rachlin, 2002; Steinbeis, 2016; Steinbeis et al., 2012, 2016). In addition, a previous intervention study that sought to enhance pro-social decision-making via a storytelling paradigm that encouraged inhibitory control saw intervention-related changes (Steinbeis & Over, 2017). The current study used one-off economic games (Dictator Game, Ultimatum Game), which are so named because the participant only completed one trial for each task, instead of repeated trials with small variations. In contrast, in the intertemporal choice and valuation tasks, participants completed more trials, but each condition (e.g., a specific combination of delayed and immediate reward values) only occurred once. This means that these tasks only provide limited information on participants' social and intertemporal decision-making. Potentially, decision-making tasks that introduce a component of learning, for example, measuring repeated value-based learning and use of effort in a pro-social setting, may prove to be a better measurement of potential training-related effects (Cutler et al., 2021; Lengersdorff et al., 2020; Lockwood et al., 2017, 2020, 2021). In addition, these types of tasks would allow further assessment of potential changes in reaction times following EF training and illuminate the fast and intuitive or slow and deliberate debate surrounding pro-social decision-making (Knoch et al., 2006; Rachlin, 2002; Rand et al., 2012; Zaki & Mitchell, 2013).

The current chapter used the percentage of delayed choices to measure intertemporal choice. While several commonly used mathematical functions are available to assess individual differences in delay discounting choice data (e.g., hyperbolic, exponential, generalized-hyperbolic, or beta-delta functions), the data did not lend itself well to fitting these functions. Many participants did not express the expected discounting patterns in their choice behavior. For example, they showed randomness in their choices by switching back after having chosen a delayed option, or they chose only 100% delayed or immediate options. Many participants with poor model fit would have had to be excluded. Instead, I decided to use a model-neutral measure, which could be extracted for every participant without violating model assumptions. Initially, I calculated the Area Under the Curve (AUC) estimate (Myerson et al. 2001), but this was so closely correlated to the percentage of delayed choices (e.g., $r > .90$) that I decided to use the most straightforward measure. I am positive that the percentage of delayed choices is the best measure I could use in this study to capture the temporal discounting preferences regarding choice. However, it is a limitation, potentially of the experimental design, that it did not lend itself to capturing the expected discounting behavior in the participants. Future studies in this age range may want to use expanded tasks, which, at minimum, include an extra block of 18 trials where both the delayed and immediate rewards are varied (Figner et al., 2010).

In conclusion, this chapter investigated the age-related changes and potential plasticity of social and intertemporal decision-making in middle childhood. While I observed the previously reported age-related increases in pro-social decisions and less steep temporal discounting, I did not observe significant relationships to EFs such as inhibitory control and working memory. However, there were some significant relationships between intertemporal choice and cognitive flexibility. Furthermore, in a randomized controlled trial of cognitive control training, targeting inhibition did not lead to short-term or long-term changes in social or intertemporal decision-making in a large sample of children aged 6-13. Thus, contrary to previous findings, I found that EFs were not a strong supporting factor in social and intertemporal decision-making in the current study.

4.6 Supplemental materials

4.6.1 Cognitive control training protocol

Participants were introduced to the training games as the ‘Treasure Game’ with the narrative that they had flown a plane, which had to crash land in the desert. To fix their plane, they were required to obtain spare parts from a sage living in a distant cave. To get to the sage, participants had to travel through 4 different worlds (i.e., forest, desert, snow, and mountains), after which they had to go back through the same worlds to return to the plane. While traveling through each world, participants could collect coins and gems, which could be used to trade for spare parts with the sage. Gems and coins were collected in the context of seven games designed to train inhibition (experimental group) and response speed (control group). The seven training games were 1) Treasure collect, 2) Mining, 3) Chest picking, 4) Conveyor belt, 5) AB Driving, 6) Hold-and-Release (HR) Driving, and 7) Forest Escape (Figure 50). Each training session entailed a combination of two games, set in a pre-assigned order at the start of training. Before starting the games, participants were presented with an option of three different caves that they could choose from to encourage engagement and a sense of agency. For both groups, sessions were recoded based on date, meaning any data logged on the same date would be grouped in the same session. Since the implementation of the games differed in terms of key presses and mechanisms tested (Table 9), I only included sessions for participants that had a minimum of two games and, for the experimental group, sessions that had at least two games with valid SSRT measures (i.e., positive SSRT values). For the control group, reaction times were included within two standard deviations of the mean reaction time per participant.

List of games
1. Treasure collector
2. Chest picking
3. AB driving
4. Forest escape
5. Mining
6. HR driving
7. Conveyor belt

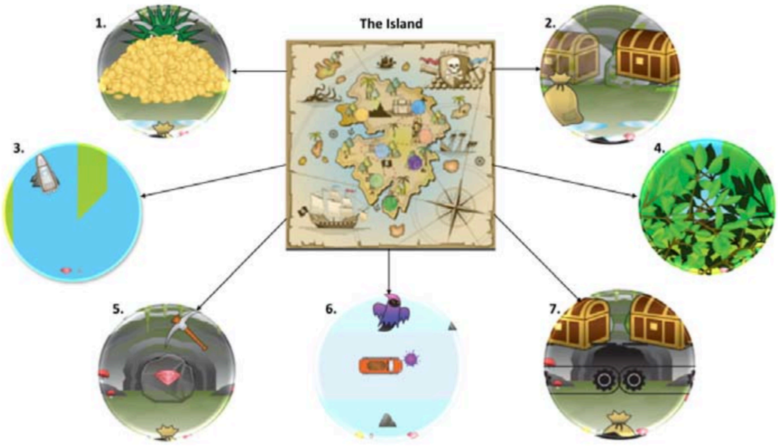


Figure 50. Seven unique training games.

Table 9. Stimulus response instructions for each game across training groups.

Game	Stimulus	Experimental Group (Response Inhibition)	Control Group (Response Speed)
Treasure collector	1. Treasure 2. Dragon	1. Press space (go) 2. Do not press space (stop)	1. Press space 2. Press space
Chest picking	1. Wobbling treasure chest on other side 2. Wobbling treasure chest on same side 3. Dragon	1. Press space to move to the other side (go) 2. Do not press space (stop) 3. Press space to move and avoid dragon (go) do not press space if dragon is on the same side (stop)	1. Press space to move to the other side 2. Do not press space 3. Press space to move to dragon
AB driving	1. Sign pointing left or right 2. Stop traffic sign	1. Left or Right Arrow key (go) 2. Do not press Left or Right arrow key (stop)	1. Press Left or Right arrow key 2. Press Left or Right arrow key
Forest escape	1. Pile of coins 2. Monster	1. Press space (go) 2. Do not press space (stop)	1. Press space 2. Press space
Mining	1. Rock 2. Gem	1. Press space (go) 2. Do not press space (stop)	1. Press space 2. Press space
HR driving	1. Ghost appears at front of car 2. Ghost appears at back of car	1. Release space (go) 2. Keep finger on space (stop)	1. Release space 2. Release space
Conveyor belt	1. Wobbling treasure chest 2. Dragon	1. Press space to change direction of the belt so that the treasure chest moves towards the bag (go) 2. Avoid the chest with the dragon by pressing space to change direction (stop)	1. Press space to change direction of the belt so that the treasure chest moves towards the bag 2. Move the chest with the dragon by pressing space to change direction

4.6.2 Imputation of missing data

The current chapter included data from three time points. Patterns of missing data are present. After Covid restrictions, behavioral testing was moved online, and the psychological battery was shortened to make independent testing at home more feasible for the participants. Due to this, the Ultimatum Game Proposer was not included in the battery (the Ultimatum Game Responder was included), and the participants who were tested online for their first timepoint hence did not have this data. In addition, there were also missing data points due to participant dropout or incomplete task batteries being completed across the time points. In this section, I report the missing data across each timepoint and the results of the imputed data.

Missing data were imputed using the MICE package in R (van Buuren & Groothuis-Oudshoorn, 2011), using Predictive Mean Matching (PMM), 50 maximum iterations, and 50 distinct datasets being generated. The imputation was based on a model generated by the Quickpred function in MICE. Decision-making measures, EF measures, intelligence, and several demographics (age, gender, training group, school) were included in the imputation. As this data was part of a more extensive training study, the entire dataset with all three time points was used for imputation. Only measures that were missing less than 70% in total were imputed. To be included, a participant must have participated in data collection for at least one time point. Reaction time measures were log-transformed before imputation.

For reference, at the pre-training timepoint, 74% of the participants had complete data, while 11% missed the Ultimatum Game Proposer data (Figure

51). And 0.04% of the participants were missing all data (they were still included, and their data was imputed because they had data in the other time points).

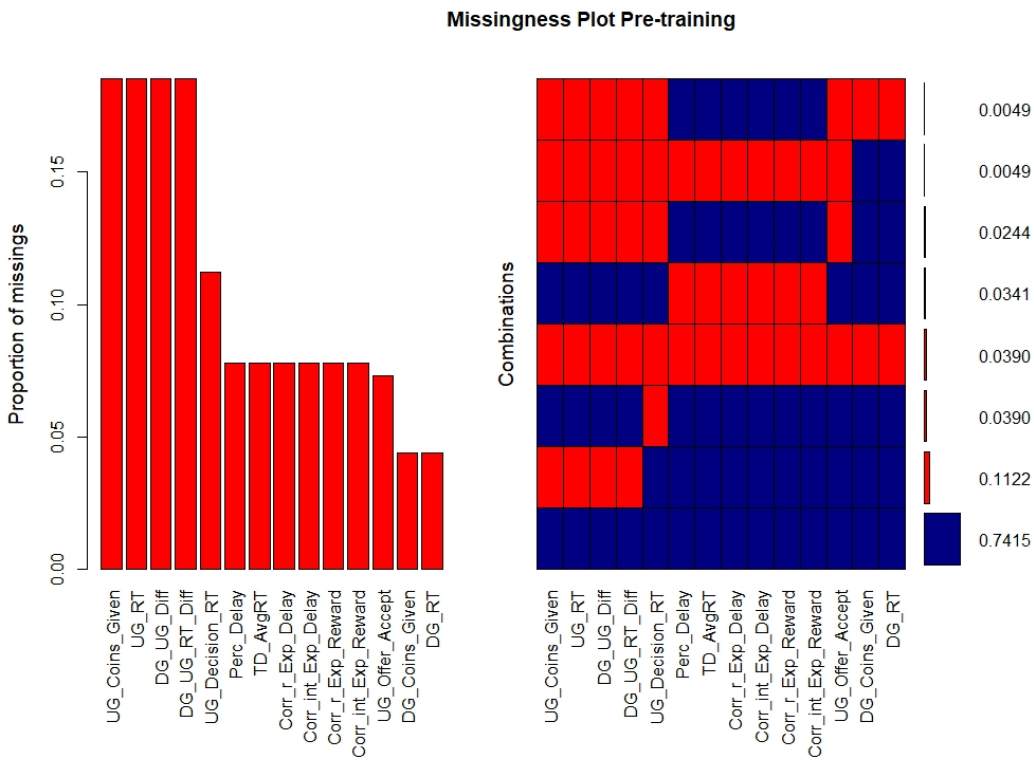


Figure 51. Pre-training missingness plot.

At the post-training time point, 38% of participants had complete data (they were tested in person and completed the whole task battery), while 22% missed the Ultimatum Game Proposer data, and 29% had complete missing data (Figure 52).

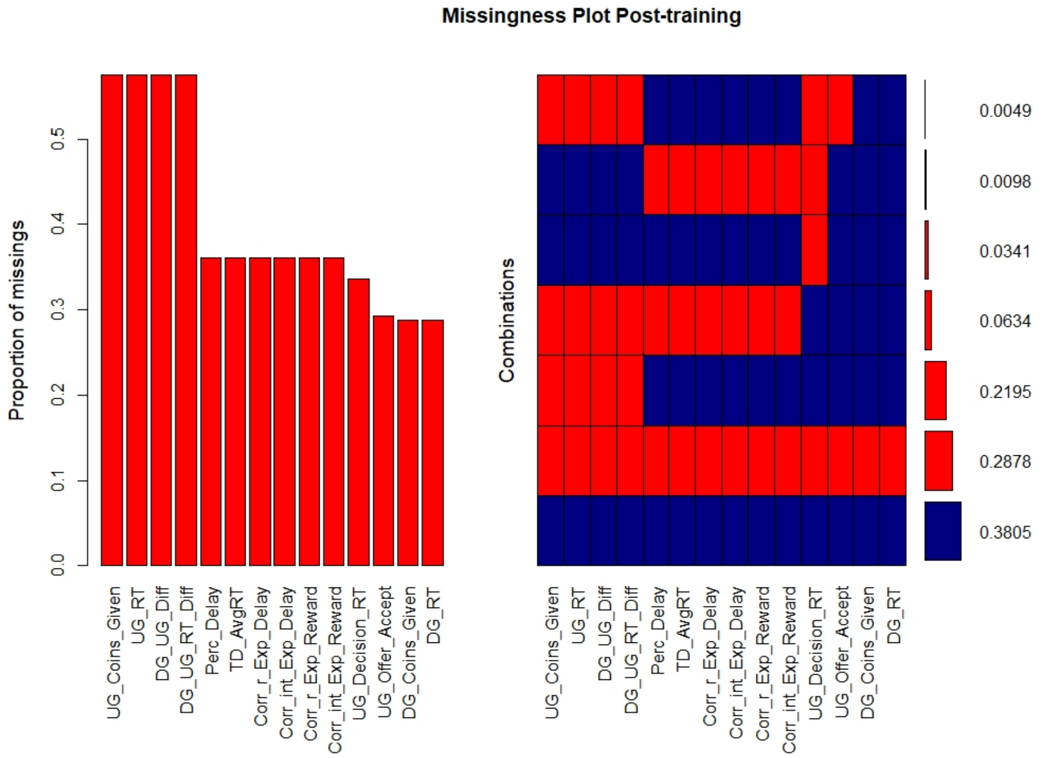


Figure 52. Post-training missingness plot.

At the one-year-follow-up time point, 72% of participants had near complete data (at the one-year-follow-up, all testing was conducted online, so all participants missed the Ultimatum Game Proposer data), while 26% of participants had complete missing data (Figure 53).

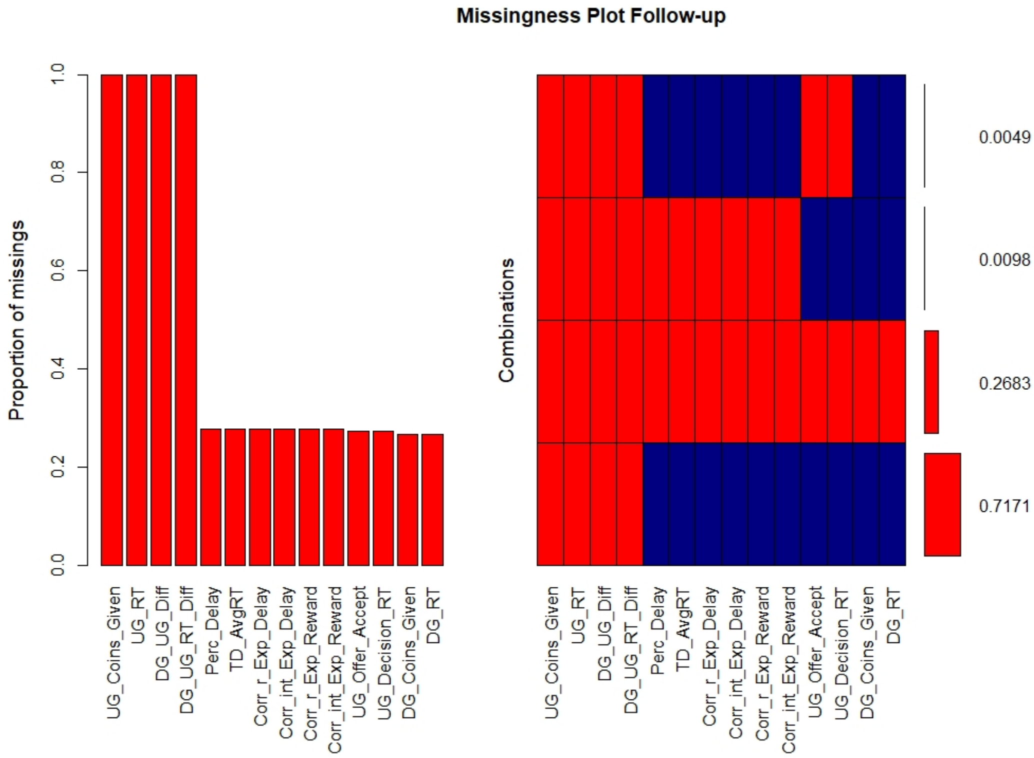


Figure 53. One-year-follow-up missing data.

After imputations, the distributions of the imputed and original datasets were inspected (Figure 54). By default, the results presented in the chapter were based on the 50 pooled datasets from the imputed data. Only the first imputed dataset was used for the factor analysis and the plots. To aid robustness, for the analyses based on the resulting decision-making factors, these factors were appended to each imputed dataset. Final analyses based on the factors report the effect sizes and errors based on all 50 imputed datasets.

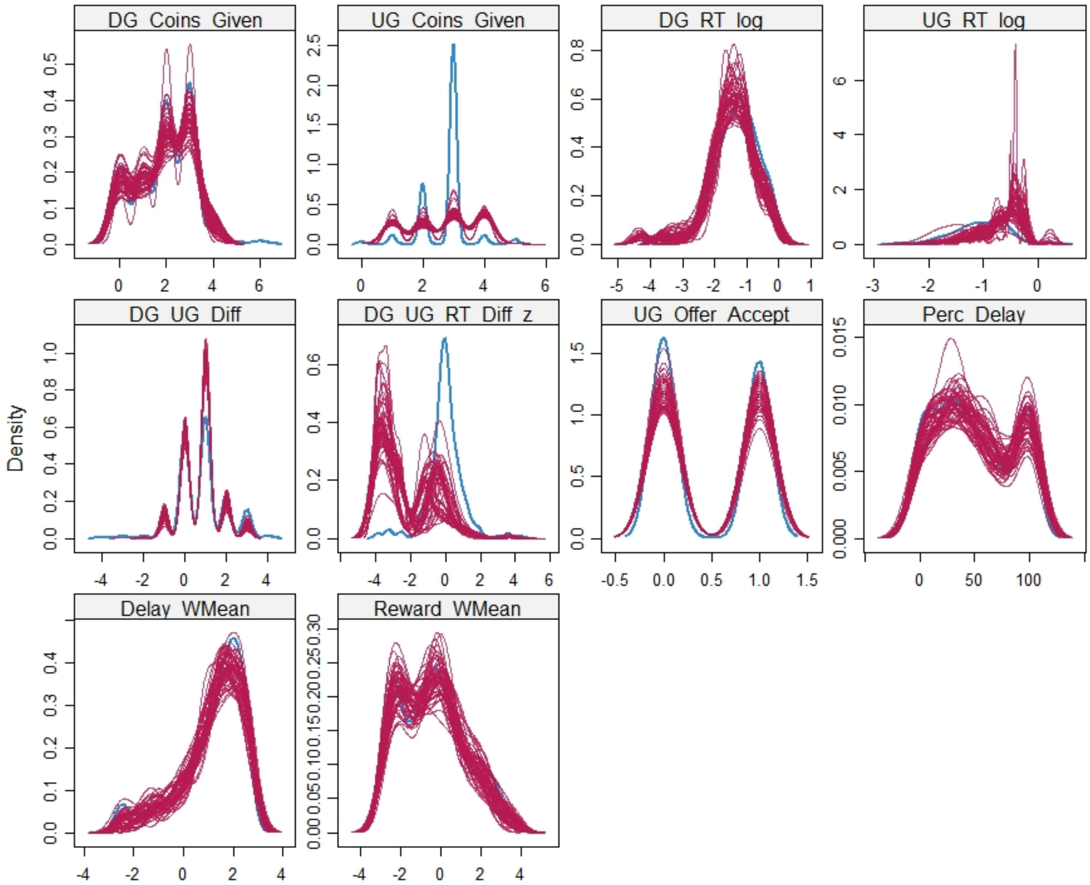


Figure 54. Distributions of the original and imputed datasets.
The original dataset is depicted in blue, and the imputed datasets in red.

4.6.3 Confirmatory Factor Analysis for executive measures

A confirmatory factor analysis (CFA) was performed using Lavaan in R (Rosseel, 2012). Full Information Maximum Likelihood (FIML) was used to deal with any missing data in the dataset. Multiple models were fit; however, the model failed to converge for most models, and some of them displayed negative variances suggesting that models were mis specified. Two models did converge: a model with a single factor encompassing all tasks and a model with three sub-factors of inhibition, shifting, and memory. There were no significant differences in model fits. As the purpose of this factor analysis was to create latent factors of the different subcomponents of EFs, the model with the three sub-factors of EFs was selected (Figure 55). Values for each individual were extracted from this for further analysis.

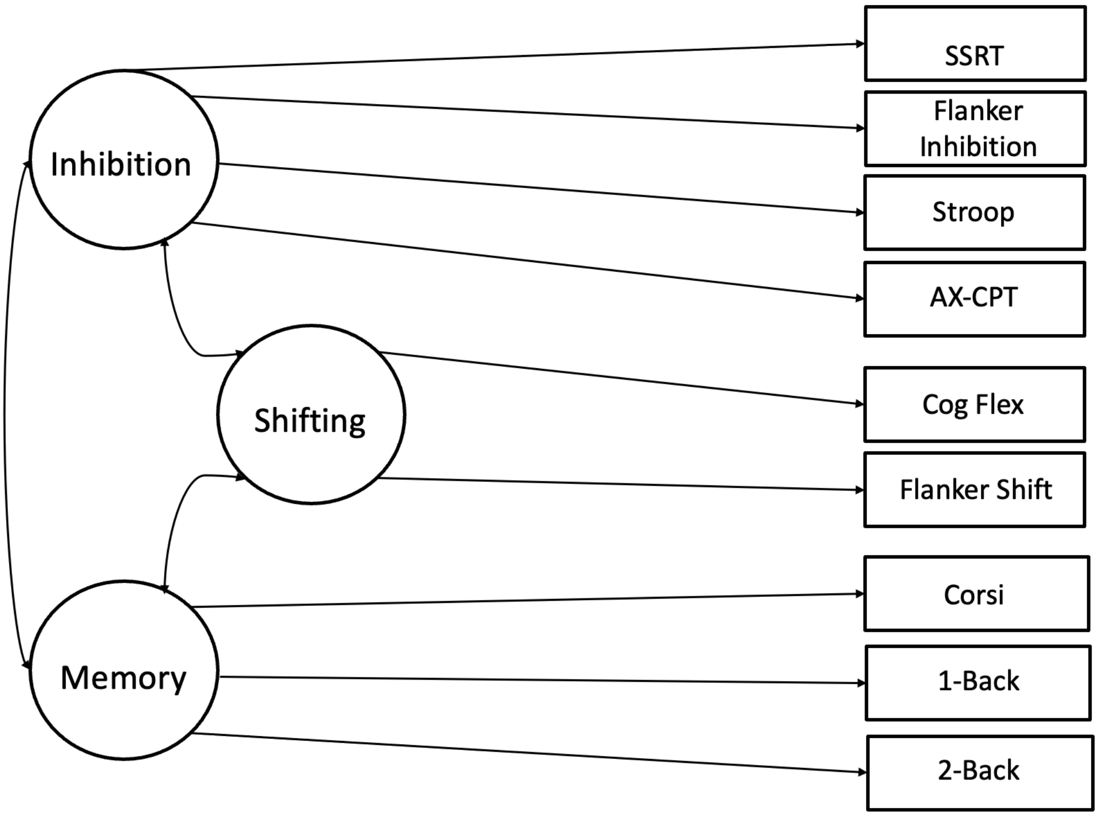


Figure 55. CFA structure for executive functions.

4.6.4 Linear valuation composite measures

To simplify the valuation measures, I created composite scores of the linear coefficients and intercepts for the delay discounting and reward valuation tasks, respectively. The linear valuation measures were created by fitting individual linear models to each participant’s valuations and extracting the coefficients and intercepts of these models. The coefficient and the intercept provide information about the individual’s valuation behavior.

For example, a more negative delay valuation coefficient indicates that a participant discounted progressive delay for the same reward (eight MUs for one to six days of delay) as progressively worse. On the other hand, a higher value for the delay valuation intercept indicates that a participant rated the initial option (eight MUs for the smallest unit of delay (delay (1 day)) as more favorable.

For the reward valuation coefficient, a more positive reward valuation coefficient indicates that a participant valued increasing immediate reward more (from two MUs, four MUs, six MUs, to eight MUs). A low value for the reward valuation intercept indicates that a participant valued the initial option (two MUs immediately) as less valuable.

When I compare this with age, I see differential relationships between the intercepts and the coefficients. For the delay valuation, there was a significant negative relationship between age and the coefficient (β = -0.11, se = 0.03, t = -3.52, $p < .001$, Figure 56a), showing that older children rated increasing delay as less favorable. On the other hand, there was a positive relationship between age and the initial rating of eight MUs for the lowest unit of delay (the intercept) (β = 0.79, se = 0.17, t = 4.75, $p < .001$, Figure 56b), showing that older children valued a high reward for less delay as more valuable, or that younger children immediately discounted a high reward for even the lowest amount of delay.

For the explicit valuations of reward, participants rated their valuations of increasing immediate rewards, starting with two MUs immediately up until eight MUs in increases of two MUs (four data points). There was a significant positive relationship between the coefficient for the valuations of reward and age (β = 0.05, se = 0.02, t = 2.19, p = .030, Figure 56c). This indicates that with age, children had a steeper increase in their valuation of increasing immediate reward. The initial valuation of two MUs immediately was not significantly correlated to age (β = -0.35, se = 0.20, t = -1.80, p = .073, Figure 56d).

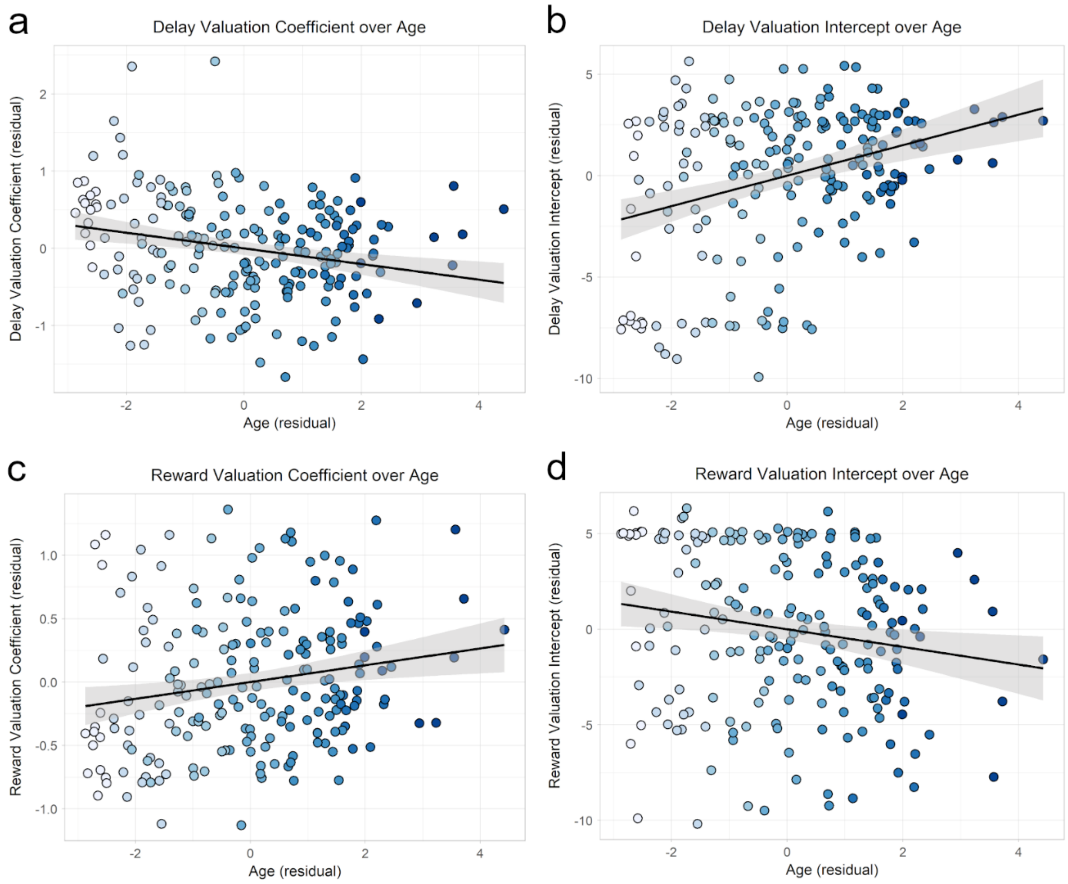


Figure 56. Linear measures derived from the delay and reward valuation data.

The top row displays the linear measures from delay valuation, reflecting the (a) steepness of delay valuation and (b) the starting valuation of the largest reward for the lowest amount of delay. The bottom row displays the measures from reward valuation, reflecting the (c) steepness of reward valuation and (d) the initial valuation of the smallest immediate reward.

The composite measures were created via weighted averaging, where the coefficients were weighed double, and the intercepts weighed once. The resulting composite scores were highly correlated to the respective coefficients and intercepts (Figure 57), (i.e., the delay valuation composite score was significantly negatively correlated to the delay valuation coefficient (β = -0.89, se = 0.16, t = -5.61, $p < .001$, Figure 57a), and significantly positively to the delay valuation intercept (β = 0.38, se = 0.01, t = 39.00, $p < .001$, Figure 57b). While the reward valuation composite score was significantly positively correlated to the reward coefficient (β = 2.61, se = 0.11, t = 23.00, $p < .001$, Figure 57c), and negatively to the intercept (β = -0.38, se = 0.004, t = -104.30, $p < .001$, Figure 57d).

The high correlations indicated that the composite scores successfully captured both the coefficient and the intercepts. The composite scores were used in further analysis of the intertemporal valuation data.

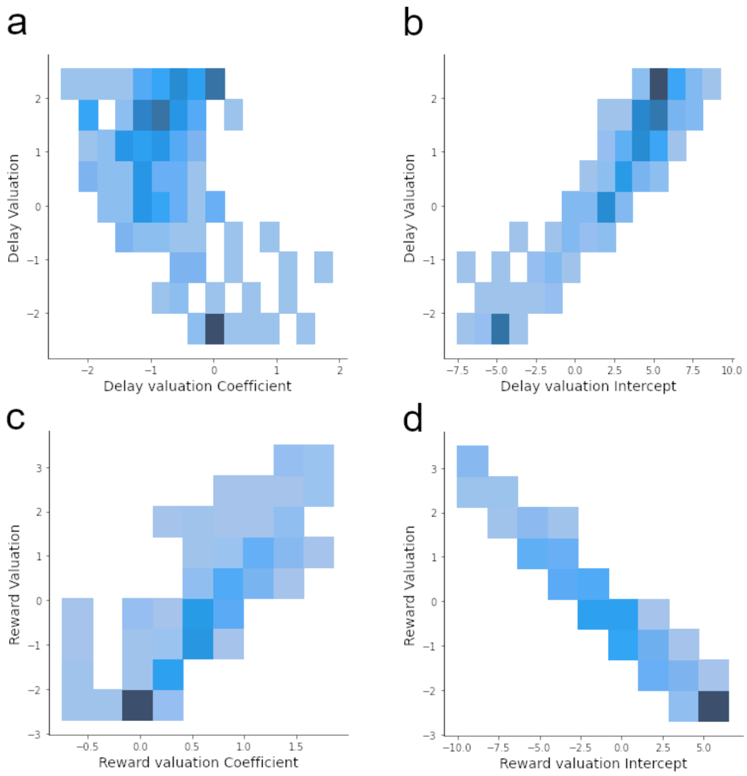


Figure 57. Correlations between the composite scores and the initial linear measures.

Correlations between the weighted averages and the respective coefficients and intercepts for delay valuation (top, a, b) and reward valuation (bottom, a, b).

4.6.5 Decision-making factor analysis

4.6.5.1 Participants

A total of 229 (120 (52%) female, 109 (48%) male) participants, with a mean age of 9.00 (SD = 1.56), and an age range from 6.03 to 13.32 years old, were recruited from a total of 20 London schools. A total of 191 participants were tested in person for the data for the factor analysis (the pre-training timepoint), and 38 participants were tested online. As the factor analysis was only conducted on the first time point (pre-training), participants were not excluded if they had not completed one valid training session. Thus, there are 24 more participants included in the factor analysis than in the current main chapter.

4.6.5.2 Statistical methods

First, the suitability of the dataset for factor analysis was assessed using the KMO and Bartlett's sphericity tests. Next, the Exploratory Factor Analysis (EFA) was conducted in R using the EFAtools package (Steiner & Grieder, 2020). To inspect the robustness, the average EFA solutions were plotted across three different methods (Least Squares, Maximum Likelihood, and Principal Axis Rotation) and 10,000 iterations. Lastly, to assess the reliability of the factors, I report McDonald's omega (sub-scale) and Cronbach's Alpha (Flora, 2020). McDonald's omega and Cronbach's alpha were calculated via the EFAtools and psych package in R.

The EFA was conducted on a single imputed dataset, but the resulting factor loadings were then appended to all 50 imputed datasets. Resulting analyses based on the factor analysis results were run by pooling over all 50 imputed datasets, offering a more robust analysis of potential associations with age, SES, and EFs.

4.6.5.3 Inter-correlations for social and intertemporal decision-making

Next, I assessed how the decision-making measures from the same task inter-correlated. First, I ran bivariate correlations. Next, I controlled these relationships for age. These tests were run on the pooled imputed datasets. The valuation measures for delay and reward were combined in a composite measure, see 4.6.4 Linear valuation composite measures.

Dictator and Ultimatum Game

First, for the Dictator Game, the number of MUs given was significantly positively correlated to the MUs given during the ultimatum game (β = 0.17, se = 0.08, t = 2.13, p = .035), this remained significant after correcting for age (β = 0.19, se = 0.08, t = 2.33, p = .022), indicating that higher offers during the dictator game were correlated to higher offers during the ultimatum game. There was a significant positive correlation between reaction time taken to distribute the MUs during the Dictator Game and the number of MUs given (β = 0.28, se = 0.10, t = 2.81, p = .006), and this remained significant after correcting for age (β = 0.24, se = 0.10, t = 2.28, p = .024). There was no significant relationship between the reaction time during the Ultimatum Game Proposer and the number of MUs given (β = 0.04, se = 0.13, t = 0.35, p = .728).

Regarding the acceptance of the unfair offer during the Ultimatum Game Responder, the time taken to decide on how to respond to the Ultimatum Game Responder was not correlated to the accepting or rejecting of the offer (β = -0.02, se = 0.04, t = -0.43, p = .671). There was no significant correlation between the probability to accept the offer and the number of MUs given during either the Dictator Game, (β = -0.10, se = 0.14, t = -0.74, p = .462). There was also no significant correlation between the probability to accept the offer and the number of MUs given during the Ultimatum Game (β = -0.18, se = 0.13, t = -1.40, p = .164).

Intertemporal decision-making

The percentage of total delayed choices was significantly positively correlated to the average reaction time (β = 0.11, se = 0.03, t = 4.40, p < .001), this remained significant after controlling for age (β = 0.11, se = 0.03, t = 4.30, p < .001), suggesting that participants that took longer to decide chose more delayed options.

The percentage of delayed choices was significantly positively related to delay valuation (β = 0.37, se = 0.04, t = 8.80, p < .001), this remained significant after controlling for age (β = 0.39, se = 0.05, t = 8.57, p < .001), indicating that a higher percentage of delayed choices was related to steeper delay depreciation. The percentage of delayed choices was also significantly positively correlated to reward valuation (β = 0.24, se = 0.06, t = 4.13, p < .001), this remained significant after controlling for age (β = 0.24, se = 0.06, t = 4.08, p < .001), indicating that a higher percentage of delayed choices was related to steeper reward valuation.

4.6.5.4 Relations between Dictator and Ultimatum Game and intertemporal decision-making

Next, I assessed how the main measures from the different decision-making measures related to each other. For the Dictator and Ultimatum Game, I focused on the number of MUs given, strategic decision-making (the difference in the MUs given), and whether the unfair UG Responder offer was accepted. I focused on the percentage of delayed choices and delay and reward valuation for intertemporal decision-making.

For the Dictator Game, a higher number of MUs given was not significantly correlated to the percentage of delayed choices ($b = 0.03$, $se = 0.07$, $t = 0.44$, $p = .659$). There was a significant positive correlation between the number of MUs given during the Dictator Game and delay valuation $b = 0.10$, $se = 0.04$, $t = 2.24$, $p = .026$), but this did not remain significant after correcting for age ($b = 0.07$, $se = 0.05$, $t = 1.54$, $p = .124$). There was no significant relation with reward valuation ($b = 0.03$, $se = 0.04$, $t = 0.66$, $p = .509$).

For the Ultimatum Game Proposer, a higher number of MUs was not significantly correlated to the percentage of delayed choices ($b = 0.02$, $se = 0.08$, $t = 0.20$, $p = .842$), or delay valuation ($b = -0.03$, $se = 0.04$, $t = -0.77$, $p = .442$), or reward valuation ($b = 0.005$, $se = 0.06$, $t = 0.08$, $p = .935$).

Strategic decision-making, or a smaller difference in the MUs given during the Dictator and Ultimatum Game, was not significantly correlated to the percentage of delayed choices ($b = -0.03$, $se = 0.09$, $t = -0.30$, $p = .761$). It was significantly correlated to delay valuation ($b = -0.14$, $se = 0.06$, $t = -2.52$, $p = .013$), but this did not remain significant after controlling for age ($b = -0.10$, $se = 0.06$, $t = -1.72$, $p = .086$). There was no significant correlation with reward valuation ($b = -0.03$, $se = 0.05$, $t = -0.56$, $p = .574$).

The probability to accept the unfair offer in the Ultimatum Game Responder was not significantly correlated to the percentage of delayed choices (b = 0.02, se = 0.04, t = 0.49, p = .628), delay valuation (b = 0.01, se = 0.02, t = 0.25, p = .803), or reward valuation (b = -0.04, se = 0.02, t = -1.58 p = .119).

4.6.5.5 Exploratory Factor Analysis

Previous studies with adults have found latent variables that explain general rational decision-making behavior. Thus I was interested if this would be true for children as well. First, I selected the most important measures of the different social and intertemporal decision-making tasks, z-scored them, and removed any highly correlated (r > 0.6) measures from the dataset. For example, the difference score between the Dictator and Ultimatum Game MUs given was left out in favor of including both the MUs given from the Dictator and Ultimatum Game to avoid high inter-correlation.

Below is an overview of the final measures in the factor analysis (Table 10).

Table 10. Decision-making measures for the Experimental Factor Analysis (EFA).

Task (with key references)	***Broad psychological domains***	***Computational constructs***	***Key individual parameters and descriptive measures***
Dictator and Ultimatum Game *(Steinbeis, 2016;*	Pro-social and strategic decision-making	MUs given away to an “other child”, strategic change in MUs given away, acceptance or rejection of an unfair	1. MUs given without social consequence (DG Game) 2. Reaction time for Giving MUs during the DG Game

Steinbeis & Over, 2017)		offer, deliberation time for unfair offer	3. MUs given with social consequence (UG Game) 4. Acceptance or rejection of unfair offer 5. Deliberation time for the unfair offer
Intertemporal decision-making (Figner et al., 2010; Steinbeis et al., 2012, 2016)	Impulsivity, future planning	Inter-temporal choice, explicit discounting of delay and reward valuation, the difference in low (less likely to discount) and high conflict (more likely to discount) choices	6. Intertemporal decision-making (percentage of delayed choices) 7. Average reaction time for choosing between immediate and delayed options 8. Delay valuation (weighted average of the integer and slope for continuous ratings of the same reward for increasing delay) 9. Reward valuation (weighted average of the integer and slope for continuous ratings of an increasing amount of immediate reward)

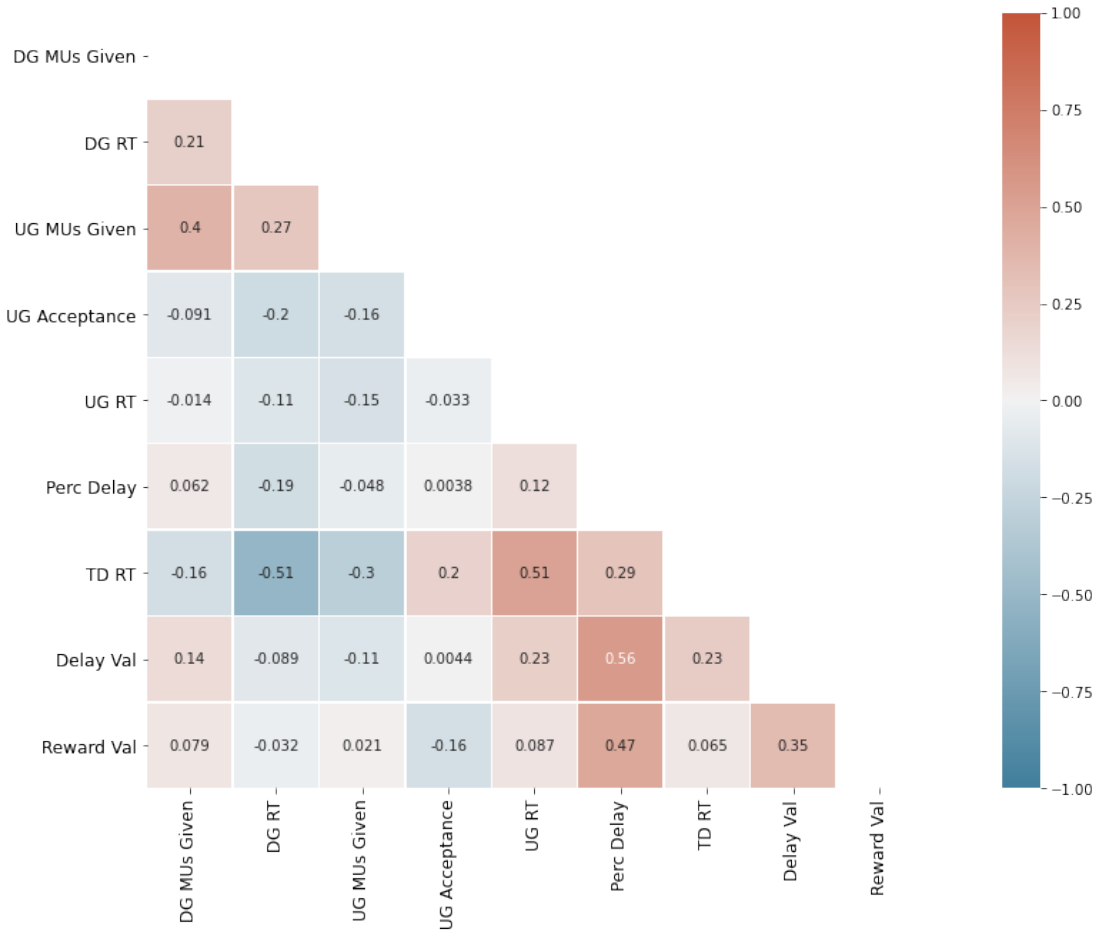


Figure 58. Correlation heatmap of the decision-making measures in the exploratory factor analysis (EFA).

The exploratory factor analysis (EFA) was conducted in R using the EFAtools package (Steiner & Grieder, 2020). First, I assessed the feasibility of the dataset for factor analysis via the Bartlett Sphericity test and the KMO test. The Bartlett Sphericity test assesses whether the dataset's variables are correlated enough that it diverges from an identity matrix (Tobias & Carlson, 1969). This test thus assesses whether a data reduction technique such as a factor analysis can be used to compress the data in a meaningful way. The Kaiser-Meyer-Olkin (KMO) test also determines how suitable data is for dimensionality reduction using factor analysis. The KMO value measures the proportion of variance among the measures that might be common (Yong & Pearce, 2013). The higher the value, the more suited the data is to factor analysis.

First, I assessed the validity of the dataset for factor analysis. The Bartlett's test of sphericity was significant ($X^2(36) = 428.82$, $p < .001$), and the KMO value at an acceptable value (KMO = 0.62). Since this value is close to 0.6, the results should be interpreted cautiously. Next, I assessed the optimal number of factors to be extracted. I used the Hull method to identify the optimal balance between model fit and the number of parameters (Lorenzo-Seva et al., 2011). Following the Hull method, the ideal number was four factors, with RMSEA and CFI as the model fits to be optimized. Then, I ran the exploratory factor analysis with four factors, using unweighted least squares as the method. I initially used an oblique rotation to assess the inter-correlation between the factors (method = "oblimin") (Yong & Pearce, 2013). The correlation was low ($r < .33$); therefore, I finally used an orthogonal rotation (varimax), which means that the resulting factors will be uncorrelated.

I found that the resulting factor structure returned four factors related to decision-making (Table 11). The model fits for the EFA were good; the CFI was at 1.00 (with values above 0.9 generally indicating good fit (Bentler, 1990), and the RMSEA was at .00 (with values below 0.05 indicating good fit (H. Kim et al., 2016).

The loadings for the first factor (F1) consisted of the total percentage of delayed choices and the weighted means for explicit delay discounting and reward valuation and could therefore be described as a less steep intertemporal discounting factor. The loadings for the second factor (F2) consisted of the UG Responder deliberation time and the average reaction time for the intertemporal choice task and could therefore be considered a deliberation factor. The third factor (F3) consisted of the Dictator Game and Ultimatum Game MUs given and could thus be described as a factor capturing social decision-making. The loadings for the final and fourth factor (F4) are harder to interpret and consisted of the probability of accepting the unfair offer during the UG Responder task and average reaction time for the intertemporal choice task, while the amount of MUs given during the UG Proposer and the average reaction time for the Dictator Game loaded negatively. I will label this factor as Choice RT & Acceptance.

Table 11. Factor loadings for the EFA.

-- Rotated Loadings ---				
	F1	F2	F3	F4
DG MUs	.108	.019	**.983**	-.138
UG Offer Accept	-.088	-.053	-.034	**.371**
UG MUs	-.038	-.106	**.365**	**-.337**
UG Decision RT	.120	**.987**	-0.36	.073
Percentage Delay	**.856**	.009	.004	.150
DG RT	-.116	-.060	.150	**-.609**
TD RT	.212	**.430**	-.099	**.690**

Delay Valuation	**.623**	.144	.068	.095
Reward Valuation	**.582**	.035	-.004	-.122
-- Model Fit --				
$X^2(6) = 4.33$, $p = .632$				
CFI = 1.00				
RMSEA [90% CI] = .00 [.00; .07]				

Since not all EFA procedures arrive at the same solution, I next performed many EFAs from different methods to provide a summary including the confidence intervals and means for the loadings across different methods (Figure 59). This plot indicates that the confidence intervals for the factor solutions were generally robust, especially for factors 1-3.

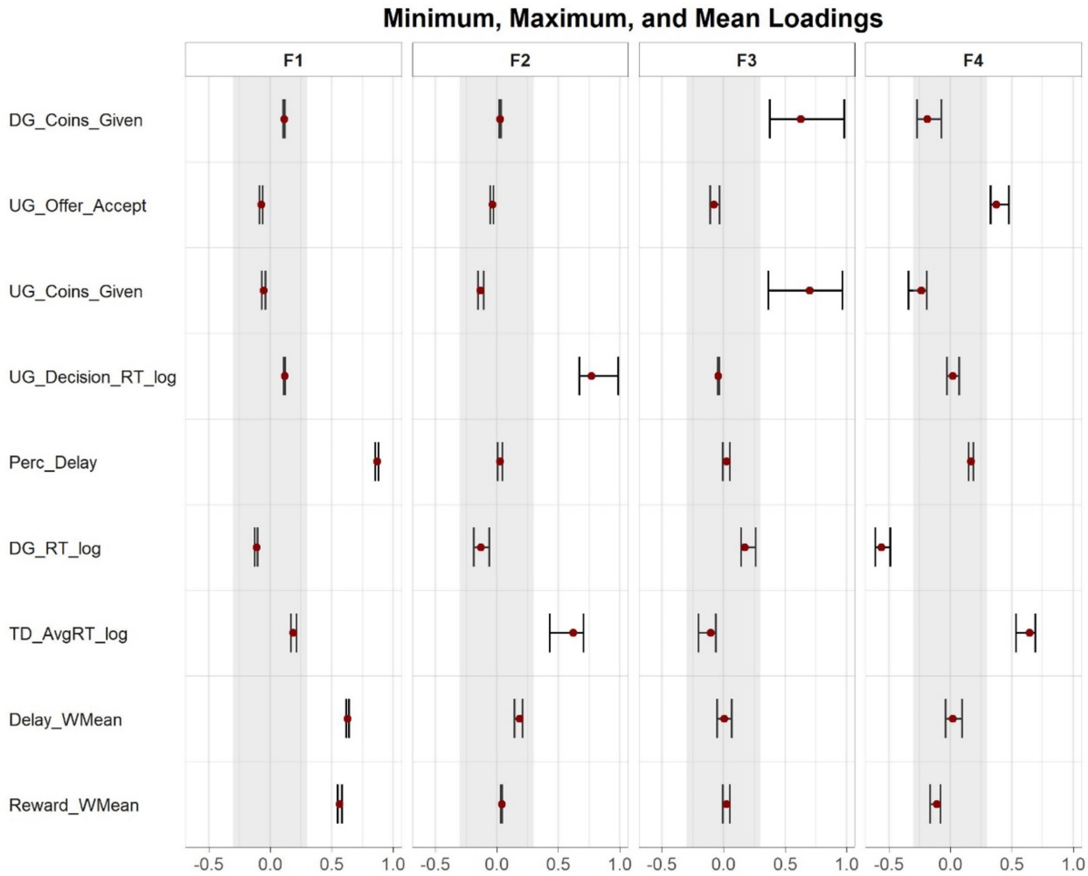


Figure 59. Average solutions and confidence intervals for the EFA across different methods (ULS, ML, PAF) and 10,000 iterations.

To assess the reliability of the factors, I report Cronbach's alpha and McDonald's omega for each of the factors (α, ω) (Revelle & Condon, 2019). Cronbach's alpha is a measure of internal consistency independent of the model used, while the omega measure depends on the model. For alpha and omega measures, a value above 0.6 is assumed to indicate sufficient consistency for a factor (Revelle & Condon, 2019).

The omega value is the most insightful in terms of internal validity. Thus, the values of both the omega and alpha measures together suggest that Factor 1 and 3 (intertemporal decision-making and social decision-making factors) have acceptable validity. In contrast, Factor 2 and 4 do not seem to have good validity (Table 12). I, therefore, continue analyses with the solutions for Factor 1 and 3 only.

Table 12. Internal validity measures for the four factor solutions.

Factors	Cronbach's Alpha (std)	McDonald's omega (sub)
Intertemporal decision-making (F1)	**0.72**	**0.60**
Deliberation (F2)	0.21	**0.68**
Social decision-making (F3)	**0.58**	**0.83**
Choice RT & Acceptance (F4)	**0.60**	0.01

4.6.5.6 Decision-making factors and executive functions

I was interested in whether the extracted decision-making factors would relate to age similarly to the separate measures and whether the factors would be associated with EFs.

First, I assessed whether the factors were correlated to age, again controlling for the testing medium. Both the intertemporal decision-making factor (β = 0.10, se = 0.05, t = 2.08, p = .039) and the social decision-making factor (β = 0.10, se = 0.04, t = 2.37, p = .019) were positively correlated to age.

Next, I assessed whether the factors were related to intelligence as measured via the WASI, which provided age-standardized measures of crystallized intelligence and matrix reasoning (Chapter 3: 3.3.3.4 Intelligence). The intertemporal decision-making factor was significantly positively correlated to matrix reasoning (β = 0.02, se = 0.01, t = 2.95, p = .004), but not to crystallized intelligence (β = 0.01, se = 0.01, t = 0.95, p = .344). The social decision-making factor was not related to either matrix reasoning (β = -0.01, se = .01, t = -1.52, p = .130) or crystallized intelligence (β = -0.0003, se = 0.01, t = -0.04, p = .964).

Finally, I assessed the potential relationships between the factors and the EF measures, controlling for age and testing medium. I used the individual EF measures also used in Chapter 3 (3.3.3 Cognitive task battery). The intertemporal decision-making factor was not correlated to any of the EF measures, either those related to working memory (Corsi: β = 0.06, se = 0.10, t = 0.62, p = .637; 1-back: β = 0.16, se = 0.89, t = 0.18, p = .859; 2-back: β = -0.08, se = 0.31, t = -0.26, p = .794), inhibition (SSRT: β = -0.03, se = 0.08, t = -0.41, p = .679; Stroop: β = 0.04, se = 0.13, t = 0.32, p = .746; Flanker inhibition: β = 0.21, se = 0.16, t = 1.34, p = .181; AX-CPT: β = -0.16, se = 0.14, t = -1.10, p = .273), or cognitive flexibility (Flanker switching: β = -0.19, se = 0.15, t = -1.29, p = .198; Dimensional switching: β = -0.27, se = 0.15, t = -1.75, p = .082).

The social decision-making factor was also not significantly correlated to any of the EF measures, either to those related to working memory (Corsi: β = -0.05, se = 0.09, t = -0.61, p = .543; 1-back: β = 0.85, se = 0.91, t = 0.94, p = .349; 2-back: β = 0.35, se = 0.29, t = 1.23, p = .222), inhibition (SSRT: β = 0.12, se = 0.07, t = 1.72, p = .086; Stroop: β = -0.05, se = 0.12, t = -0.37, p = .711; Flanker inhibition: β = -0.25, se = 0.15, t = -1.65, p = .102; AX-CPT: β = 0.08, se = 0.13, t = 0.58, p = .564), or cognitive flexibility (Flanker switching: β = 0.02, se = 0.14, t = 0.11, p = .914; Dimensional switching: β = 0.16, se = 0.14, t = 1.19, p = .234).

Thus, both factors were positively related to age, and the intertemporal decision-making factor was correlated to higher matrix reasoning, but there were no significant associations between the decision-making factors and any of the EF measures.

Chapter 5. General discussion

Part of Chapter 5; section 5.9, has been previously published in a review paper:

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a science of effective cognitive training. Current Directions in Psychological Science, 29(6), 531-537.

5.1 Summary of experimental chapters

This thesis aimed to investigate the different strategies children use when making decisions and how these might depend on underlying neurocognitive underpinnings such as cortical thickness and executive functions (EFs).

In the first experimental chapter, I researched whether using a sequential decision-making task that rewarded the use of more cognitively effortful decision-making, 85 children aged 5-11 years old would display model-based decision-making. I found robust behavioral and computational markers of model-based decision-making in this age group, in contrast to previous developmental studies that found no markers of model-based decision-making in children before 12. Next, I compared the children's behavior to a reference sample of 24 adults aged 18-35. I found that adults displayed higher levels of model-based decision-making and the stake effect (i.e., successful metacontrol by prioritizing model-based decision-making for high-stake trials over low-stake trials). In contrast, children did not show a robust stake effect and thus showed no consistent metacontrol as a group.

In the second experimental chapter, I further explored the neurocognitive underpinnings of model-based decision-making in childhood and its metacontrol in a new sample of 69 children aged 6-13. In this chapter, I replicate the findings of the previous chapter, in that there were both computational and behavioral markers of model-based decision-making and an absence of a group effect of metacontrol for children. Next, I sought to link model-based decision-making to measures of EFs across several domains (inhibition, cognitive flexibility, working memory) and intelligence. In addition, I investigated whether individual differences in cortical thickness were related to model-based decision-making. I found no significant relationships between EFs and cortical thickness to model-based decision-making in 6–13-year-old children for both these analyses. Finally, I also sought to link EFs and cortical thickness to individual differences in metacontrol. Surprisingly, higher metacontrol was linked to worse performance on incongruent trials on a Flanker inhibition task. Specifically, individuals with higher metacontrol seemed to prioritize performance on the congruent over the incongruent trials. In addition, in a whole-brain analysis, I found that increased cortical thickness in several brain regions linked to memory and contextual learning was linked to higher metacontrol. Furthermore, in an ROI analysis, higher cortical thickness of the bilateral DLPFC was related to higher metacontrol. Thus, metacontrol was linked to performance on an EF task and individual differences in cortical thickness.

In the third and final experimental chapter, I investigated the potential relationships between EFs and social and intertemporal decision-making in 6–13-year-old children, both from a correlational approach and with a training paradigm. Specifically, I researched whether training EFs by targeting inhibition led to short- or long-term training-related increases in pro-social and intertemporal decision-making. First, I found several relationships between EFs and decision-making measures; however, most of these did not survive correction for age. The only significant relations to survive were a positive relationship between the cognitive flexibility factor and a higher percentage of delayed choices and steeper delay (de)valuation. To test the effectiveness of the inhibition response training paradigm, I reported Bayesian evidence for the null effect (no effect of training). I found strong evidence for the null, and thus there seem to be no short-term or long-term effects of inhibition response training on social and intertemporal decision-making.

5.2 On the current and previous contrasting findings of model-based decision-making in childhood

In the first chapter, I found evidence of a sophisticated and planning-based approach to value-based decision-making for 5–11-year-old children. Specifically, I found behavioral and computational markers of model-based decision-making, in contrast to previous work with children suggesting that model-based decision-making would only become apparent later in adolescence (Decker et al., 2016).

Daw and colleagues originally developed the two-step task in 2011 to investigate the differential contributions of two types of decision-making: habitual and goal-directed (Daw et al., 2011). While this task gave insight into the duality of these two types of human decision-making, the foundations of the two distinct styles differed in cognitive cost to employ (Daw, 2018; Daw et al., 2011). Therefore, it would make sense that participants would only engage in the more cognitively effortful strategy if there was also a pay-off involved (Kool et al., 2016, 2017; Kool & Botvinick, 2014). However, in the original version of the two-step task, the display of model-based decision-making was spontaneous and was present even in the absence of increased rewards for efforts in healthy adults (note: mainly young adult university students, which may also not be a representative sample). Later developmental studies that investigated model-free and model-based decision-making in developmental samples tended to use the same task, although adapted in a child-friendly way in terms of stimuli and task narrative (Decker et al., 2016; Potter et al., 2017). However, these studies did not observe clear behavioral or computational markers of model-based decision-making in children before the age of 12, leading to the interpretation that model-based decision-making may be a uniquely late-developing skill (Decker et al., 2016; Potter et al., 2017).

However, literature on effort avoidance in development suggests that children may be susceptible to increases in effort and only willing to engage in more effortful tasks for substantially larger pay-offs (Chevalier, 2018; Ganesan & Steinbeis, 2021; Niebaum et al., 2019). Thus, these studies were ultimately unable to determine whether the absence of the display of model-based decision-making was due to a developmental limitation or whether children in these studies did not find it worthwhile to engage in a more effortful decision-making strategy. In Chapter 2, I present evidence for markers of model-based decision-making in children younger than 12 years old.

There were critical task-related differences between this work and previous work concerning the task design. As it is a relatively common finding in developmental literature for complex cognitive abilities to shift to lower ages with reductions in task complexity (Scott & Baillargeon, 2017), this could be the case here as well. This thesis used a recently developed task that made engaging in model-based decision-making more rewarding in that the degree of model-based decision-making displayed was also correlated to the number of rewards a participant would win during the task (Kool et al., 2016). This was mainly achieved through five major changes to the task paradigm, which were:

1. a switch to deterministic instead of stochastic transitions,
2. an increased rate of change for the drifting reward rates during the task,
3. a point-based system of reward rather than binary wins/losses,
4. coupled with a broader range in scale of rewards (from 0-9 rather than a probability of reward from 25% to 75%), and
5. a simplified task structure, meaning there was only one actual decision for participants (i.e., they only needed to choose a rocket to travel to a planet and did not need to choose between aliens while on the planet).

These changes either reduced the complexity of the original task (1, 5), increased the amount of information available to participants (3, 4), or increased the speed with which rewards change, meaning that making the right decision *all the time* became more important (2). While these changes seem substantial, the essence of the model-free and model-based dichotomy is preserved; participants either use the underlying structure of the task to plan their next decision, or they respond habitually without planning through the task (Kool et al., 2016, 2017).

In addition, a recent study suggested that reward-related learning across the lifespan remains relatively stable and that instead, action biases decrease, and successful learning from punishments increases with age (Pauli et al., 2022). Previous versions of the two-step task used probabilistic reward; participants either won a reward or not. In contrast, in the current task, participants received rewards across a point scale. Thus, it is possible that this point-based reward system proved to be easier to learn from for the children, while it may be differential learning from losses that hampered developmental samples previously in employing model-based decision-making.

To summarize, the changes in the current paradigm made the task less complex and reduced the uncertainty. This made choosing the right way to engage with the task (applying the task structure) easier to do and more rewarding. This could have likely explained why children now engage in model-based decision-making. Thus, I mainly attribute the differences in the current thesis and previous work due to task differences.

5.3 On the absence of a link between executive functions and model-based decision-making in childhood

In the first two sections of the introduction, I mentioned that essential concepts in decision-making research, such as "bounded rationality" and "The Law of Effect", link optimal and rational decision-making to cognitive abilities. In the Law of Effect, the ability to associate past actions and outcomes in certain situations is supported by the ability to remember the previous circumstances and outcomes of making a decision and searching through alternative options. According to the concept of bounded rationality, our rational decision-making is bounded by the human mind's capacity to only entertain a few alternatives at a time.

In Chapter 3, I investigated the potential relationship between computationally costly model-based decision-making and its metacontrol to EFs, reflecting working memory, cognitive flexibility, inhibition, and intelligence. While previous studies linked working memory (Otto, Raio, et al., 2013), cognitive control (Otto et al., 2015), and intelligence in the form of fluid reasoning (Potter et al., 2017) to model-based decision-making, I did not find such relationships in the current sample of 6-13-year-old children. Several factors may explain this absence. In a previous paper, working memory ability was related to model-based decision-making in adults, but only indirectly, in protecting against the influence of stress (Otto, Raio, et al., 2013). A relationship between model-based decision-making and working memory was found in a paper with a sample aged 9-25 years old. However, the working memory measure for many participants hit ceiling levels (Potter et al., 2017). While ceiling effects can lead to true relationships between performance on two tasks being obscured, if a high level of working memory is indeed related to a higher display of model-based decision-making, this relationship can still be significant. In this study, there was no ceiling to the working memory span measure, and I did see a correlation between model-based decision-making and working memory that was no longer significant after correcting for multiple comparisons. Thus, it may be that the use of a more sensitive working memory measure led to the absence of this relationship and that in the previous study, while the previous relationship between the working memory measure at ceiling might have reflected other abilities.

The previous study also found a significant relationship between fluid reasoning and model-based decision-making (Potter et al., 2017); however, I did not replicate that finding in the current study. Although, the previous research reported using raw scores, and every participant completed the first 32 items of the WASI matrix reasoning task, regardless of age (Potter et al., 2017). In contrast, the current study used age-standardized scores, and the participants completed the maximum number of puzzles they were able to, which meant the task was stopped if they reached three incorrect items in a row, limited by their age according to the WASI guidelines (participants younger than nine did not complete the last five puzzles, even if they were able, with the maximum number of items that can be completed being 35) (Wechsler, 2011). Therefore, the difference in using raw and age-standardized scores and a cap on the maximum number of items participants could complete in the current study may explain this absence.

Another previous study found a relationship between inhibition as measured via the Stroop task and cognitive control measured via the AX-CPT task and model-based decision-making (Otto et al., 2015), and I did not find a significant link between model-based decision-making and the AX-CPT task in the current study. However, there were several important differences between the previous and current studies. First, the previous study used an adult sample, and the current study a developmental sample; secondly, the previous study used the classical visual Stroop paradigm, and the present study a visual-auditory version of Stroop; lastly, the AX-CPT task used in the previous study had more trials and a different composition of AX and AY, BX and BY trials. Therefore, as the ages between the samples were different, and the tasks themselves had small differences, this could clarify the absence of these relationships in the current study.

The absence of the relationship between working memory and model-based decision-making in this thesis could be due to a specific aspect of the current two-step task. By simplifying the task design, for example, by reducing the second stage of the task (the planet stage) to a single choice (press spacebar to receive reward) rather than an additional choice (e.g., in Daw et al. 2011, participants had to choose between two aliens on each planet, each with a different drifting reward rate), the working memory load needed to make good decisions is reduced. Instead, the focus of the task is almost solely on integrating the transition structure in the first stage, when participants choose a rocket to travel to a planet. This simplification thus allows assessing whether participants are capable of using the transition structure in the task to plan through the task structure, as the surrounding “noise” of the original task is reduced. Thus, the working memory load is limited to remembering a single step, i.e., the choice of which rocket to use to travel to the desired planet. It should be noted that even with this simplification, doing the task is still not easy, especially with the added time pressure and the rapidly changing reward distributions. In addition, participants need to monitor and remember the current state of the planets depending on the drifting reward rates. However, the working memory load is lessened. This could thus explain the absence of a link between working memory and model-based decision-making in the current sample.

To my knowledge, the current work is the first to look at the relationships between EFs and model-based decision-making in 6–13-year-old children. There is the possibility that in this age range, the cognitive functions are still separate, whereas, by adulthood, cognitive processes are thought to become more integrated (Luna et al., 2001; Luna & Sweeney, 2004; Montez et al., 2017). For example, a study with participants between the ages of 11 and 37 observed that age-related changes in executive cognitive ability were related to increased neural network functional organization, where brain regions recruited for cognitive control showed more functionally integrated brain activity with age (Stevens et al., 2009). Thus, previous relationships between EFs, such as working memory and cognitive control in adulthood, may not yet hold in childhood. However, as previously mentioned, it could also be true that the reduced complexity of the current two-step task design reduces the need to rely on other EFs. Alternatively, changes in the EF task paradigms across studies led to slight variations regarding the exact domain being measured. Future studies investigating model-based decision-making across a more extensive age range could illuminate when EFs begin to predict model-based decision-making and whether this differs across different two-step paradigms.

Finally, to draw meaningful and reliable inferences based on performance on cognitive tasks, the test-retest reliability of such tasks should be considered. Poor test-retest reliability can have several implications such as problems with validity; the interpretation of results; and the generalization of results. Validity problems can call into question the ability of the task to measure the construct it was designed to assess. Interpretation problems make it difficult to interpret individual differences in performances as real differences in ability or cognition, or whether they reflect other factors such as mood and fatigue. Finally, poor test-retest reliability can lead to problems with generalizing results and inconsistencies between studies.

In the current thesis, several cognitive tasks have previously been reported to have high test-retest reliability, such as the dimension switching task ($r = 0.77$) (Paap & Oliver, 2016), the AX-CPT ($r = 0.73$) (Weafer et al., 2013), and the N-back task ($r = 0.70$) (White et al., 2018) and some moderate, such as the Stop-Signal Task ($r = .65$) (Weafer et al., 2013), the Flanker task ($r = 0.64$) (Paap & Oliver, 2016), and the Stroop Task ($r = .50$-0.80) (Penner et al., 2012), and some low reliability, such as the Corsi (WM Span) ($r = 0.28$) (White et al., 2018). Thus, while most cognitive tasks used in this thesis offer acceptable levels of test-retest reliability based on previous studies, interpretations on the forward working memory span as measured via the Corsi, should be interpreted with caution.

5.4 On the model-free and model-based dichotomy

While the dissociation between model-free and model-based decision-making has been widely studied and supported (Bolenz et al., 2019; Bolenz & Eppinger, 2021; Doll et al., 2015; Gläscher et al., 2010; Kool et al., 2016, 2017; Otto et al., 2015; Otto, Gershman, et al., 2013; Patzelt et al., 2019), recent studies suggest that this dichotomy might be oversimplified, as well as potentially under-estimating 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 propose a more nuanced view of 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). In short, the successor representation model is an intermediate between model-free and model-based systems, which balances the efficiency of model-free decision-making with the flexibility of model-based decision-making by storing partially computed action values which are predictions about future events (Momennejad et al., 2017). This model is appealing because, from behavioral data, we see that humans already express a hybrid form of decision-making between model-free and model-based decision-making. Thus, a model that combines the best of both may be able to more closely approximate human decision-making.

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 the current work, the dichotomy has aided in understanding whether children aged 5-11 years old could apply an underlying transitional structure to their decisions and is a valuable contribution to the field in including a more comprehensive range of developmental research.

5.5 Interpreting the relationship between metacontrol and

executive functions

In Chapter 3, I researched the neurocognitive underpinnings of model-based decision-making and metacontrol. While I did not observe the same relationships between EFs and intelligence with model-based decision-making as in previous studies, I did see a relationship between performance on an inhibition task and metacontrol, although not in the expected direction. Instead, a prioritization of performance on the congruent trials in a Flanker task, rather than the incongruent trials, was linked to a higher degree of metacontrol (for an overview of the task, see 3.3.3.1 Inhibition and Figure 21).

I believe there could be two reasons for this relationship. First, if children are highly susceptible to effort, especially in the absence of increased reward (Chevalier, 2018; Niebaum et al., 2019; Niebaum, Jesse C., Chevalier, N., Guild, R.M., Munakata, 2020), then the absence of increased reward for the Flanker inhibition task's incongruent trials over the congruent trials could explain why this is mirrored in prioritization of the high-stake trials in the two-step task. Thus, as higher metacontrol is linked to the avoidance of the more difficult trials on the Flanker task, this could indicate that this reflects proactive control over responses and when to exert effort. Previous studies have shown that younger children especially seem unable to avoid more effortful tasks or conditions, which may reflect that they do not yet have the cognitive control necessary to avoid these higher-effort sections (Ganesan & Steinbeis, 2021; Niebaum et al., 2019; Niebaum, Jesse C., Chevalier, N., Guild, R.M., Munakata, 2020).

Therefore, cognitive control may be the force through which cognitive effort can be selectively exerted (Shenhav et al., 2017). It has been argued elsewhere that the development of cognitive control during childhood also requires that children coordinate available control strategies more flexibly as they age (Chevalier, 2015). This ensures dynamic adjustment of control engagement to match moment-to-moment fluctuations in contextual demands. Studies on the development of effort-related decision-making have shown that between the ages of 6-12, children become increasingly able to coordinate their behavior to contextual demands on their cognitive control (Niebaum et al., 2019). Supporting this view, I observed that a higher degree of metacontrol was also linked to a higher display of model-based decision-making overall (3.4.1 Markers of model-based decision-making and metacontrol); thus, this suggests that participants with higher metacontrol can selectively increase their model-based decision-making when it is worth their while.

Second, the alternative explanation is that higher metacontrol is related to worse inhibitory performance in childhood. However, when I split up the metacontrol and Flanker inhibition measures, I saw that this seems to reflect different prioritizations: participants with higher model-based decision-making during the high-stake trials performed better on the congruent trials. In comparison, participants with higher model-based decision-making during the low-stake trials performed better in the incongruent trials. Thus, rather than the relationship between metacontrol and Flanker inhibition being driven by an impairment in the incongruent trials, it suggests that it is driven by better performance in the congruent trials. Therefore, the interpretation in the previous paragraph seems more sensible: higher metacontrol is reflected in more selective prioritization in the Flanker inhibition task. In sum, higher metacontrol seems to be reflected in the ability to exert effort selectively in an inhibitory control EF task.

Finally, while the susceptibility to effort in children offers a plausible explanation of the current results, there is conflicting evidence regarding effort in relation to performance in childhood and adolescence. Some studies have found that children choose to expend more effort, while others have found them to be highly avoidant of it (Chevalier, 2018; Niebaum et al., 2019). For example, children have been often found to spend more effort on exploration, even when they are familiar with the optimal decision trajectory for exploitation (Meder et al., 2021). Thus, while in the current thesis, children aged 6-11 seemed to be able to selectively choose when to exert effort, further confirmation of this in future work could provide more substance to this hypothesis.

5.6 Metacontrol and cortical thickness

Chapter 3 investigated the neurocognitive correlates of model-based decision-making and metacontrol. To this end, I researched their relationship with EFs and cortical thickness. While I did not find robust effects for model-based decision-making for neither EFs nor cortical thickness, for metacontrol, I found several. Two sets of cortical thickness analyses were included. First, I ran exploratory whole-brain analyses controlled for age and sex. Second, I ran an ROI analysis on the DLPFC.

Two clusters were significant for the effect of metacontrol in the whole-brain cortical thickness analysis. They survived family-wise error (FWE) correction and were located in the left temporal and right parietal lobes. Clusters were mapped to anatomical regions using the Desikan-Killiany atlas (Desikan et al., 2006). The cluster in the left temporal lobe spanned areas involved with memory (Druzgal & D'Esposito, 2001; Jessen et al., 2006; Mion et al., 2010; Rodrigue & Raz, 2004), as well as contextual learning (Aminoff et al., 2007; Coutureau & di Scala, 2009; X. Peng & Burwell, 2021). Contextual-based learning is relevant as metacontrol in the current study represents the ability to increase computationally effortful performance when beneficial selectively. The right cluster spanned areas that have previously been linked to working memory (Koenigs et al., 2009), cognitive control (Loose et al., 2017), and planning (Randerath et al., 2017). Thus, these clusters span brain regions previously implicated in cognitive abilities relevant to metacontrol. In addition, the previous link between the superior parietal cortex with cognitive control and planning is relevant, as active prioritization of when to employ model-based decision-making across contexts relies on being able to control when to use which decision-making strategy and selectively switching between them based on context.

Using an ROI analysis, I found that the cortical thickness of the bilateral DLPFC was positively related to increased metacontrol, a brain region previously found to be involved in cognitive control and computationally effortful decision-making strategies (Beierholm et al., 2011; Cremer et al., 2021; Doll et al., 2015; Gläscher et al., 2010; S. W. Lee et al., 2014; O'Doherty et al., 2015; Smittenaar et al., 2013). Surprisingly, cortical thickness of the DLPFC was not linked to model-based decision-making, and no significant clusters remained for model-based decision-making for the exploratory whole-brain analysis. Previous studies linking the DLPFC to model-based decision-making often focused on adult functional activity, and the current study used brain anatomical measures, which may explain these differences. However, structural differences in brain anatomy may be a preferred method of assessing individual differences as opposed to functional activity, as it supplies a more stable measure, less easily influenced by daily variations in attention (Fjell et al., 2015; Karama et al., 2011).

While the sample used in Chapter 3 (N = 44) is a relatively sizeable independent childhood sample for MRI data and the largest to date for MRI studies investigating model-based decision-making cited in this thesis (e.g., Smittenaar et al. 2013: N = 25; Glascher et al.: N = 20; Daw et al. 2011: N = 17), a recent study investigating brain and behavior associations have suggested that a sample of less than a thousand participants might be suffering from being underpowered and having inflated effect sizes (Marek et al., 2022). Thus, while the relationship between individual differences in brain anatomy and behavior found in this study is interesting, this is shown in a relatively small sample across a developmental age range. Initially, I had planned to collect more MRI data. Unfortunately, due to the onset of Covid-19, I could not continue MRI scanning.

However, several methods were employed in the current study to aid the reproducibility and robustness of the current findings. The MRI data used in the current study is of high resolution (3T) and high quality. Quality of the scans was assured before analysis, where the MRI data was thoroughly inspected and cleaned in multiple iterations, and the data were smoothed with a Gaussian kernel and registered to a standard template. For the whole-brain assessment, the cluster-defining threshold was set to $p < .01$. In addition, I limited the ROI analysis to a brain area that has been implicated previously in the context of model-based decision-making and metacontrol (Daw et al., 2005, 2011; Gläscher et al., 2010; Glascher et al., 2012; Smittenaar et al., 2013). Lastly, I limit the brain-phenotype associations to direct behavioral measures completed by participants as opposed to indirect measurements (e.g., questionnaires) (Marek et al., 2022). Therefore, the current work is one of the largest (developmental) samples investigating the brain correlates of model-based decision-making and metacontrol and following good standards for structural MRI analysis. While many of the arguments put forward in the Marek et al. study are valid and should be incorporated to the best ability, independent developmental studies face difficulty in collecting large samples of MRI-based data. Future work would ideally reproduce and expand on these findings in more extensive samples so that the robustness of the current results may be further assessed.

5.7 Social and intertemporal decision-making and executive functions

Chapter 4 investigated how social and intertemporal decision-making related to EFs. Specifically, I used EF factors based on the measures included in Chapter 3 (3.3.3 Cognitive task battery) to reduce the task-related variance and create a construct measure. The three subdomains of EFs were determined a-priori (i.e., inhibition, cognitive flexibility, and working memory), and the individual measures were loaded onto these measures.

Initially, I created a single decision-making factor of the social and intertemporal decision-making measures to determine whether these measures would load onto a single factor (see 4.6.5 Decision-making factor analysis for the details). However, I found no general decision-making factor, such as the one found in Moutoussis and colleagues (Moutoussis et al., 2021). Instead, I found that the social and intertemporal decision-making measures could independently be condensed into a factor. As there was no general measure, I continued with the independent decision-making measures and related these to the EFs' latent factors.

First, I investigated the relationship between social and intertemporal decision-making and age. I replicated some of the previously reported age-related effects (Matsumoto et al., 2016); for example, pro-social decision-making, as determined via the Dictator Game, increased with age. However, strategic decision-making did not increase with age but decreased, contrary to previous studies (Steinbeis et al., 2012). This was due to younger children increasing their offers in the Ultimatum Game Proposer, while the older children had more consistent offers. Contrary to previous work (Blake et al., 2014; Blake & McAuliffe, 2011), I did not see age-related changes in the probability of accepting the unfair offer of the Ultimatum Game Responder. Instead, I observed that older children had longer reaction times for all social decision-making measures.

In the Dictator Game, I found that longer reaction times to distribute the MUs led to more MUs given away or more pro-social behavior (see 4.6.5.3 Inter-correlations for social and intertemporal decision-making); this remained significant after controlling for age. Thus, this suggests that more protracted deliberation led to more pro-social decision making. In addition, longer reaction times when deciding between the immediate and delayed choices were correlated to a higher percentage of delayed choices, which also remained significant after controlling for age. Previous studies found that people may make pro-social decisions slower (Knoch et al., 2006; Krajbich et al., 2015; Rachlin, 2002), while people may make faster decisions when they relate to themselves (Lockwood et al., 2016). In addition, longer reaction times have previously been linked to making more delayed choices (Weber & Huettel, 2008). This finding thus seems to be in concordance with previous work.

For intertemporal decision-making, surprisingly, there was no relationship between age and the percentage of delayed choices. This thus seems to counter the theory that children become less susceptible to temporal discounting with age (Green et al., 1999; Steinbeis et al., 2016). However, children valued increases in delay for the same reward as progressively worse with age. There were also no age-related effects in reward valuation, whereas I expected this to increase with age (Veselic et al., 2021). Thus, I observed some common age-related changes in social and intertemporal decision-making. However, I still observed substantial individual differences in the social and intertemporal decision-making measures, so this leaves the question if other underlying factors may predict these differences. This brings us back to the concept of bounded rationality.

There are previously reported relationships between pro-social and intertemporal decision-making and EFs (Basile & Toplak, 2015; Bickel et al., 2014; Hughes & Ensor, 2010; Kenny et al., 2016; Mellis et al., 2019; Steinbeis et al., 2012; Wesley & Bickel, 2014). As mentioned in the introduction, increased pro-social decision-making has previously been linked to better cognitive control in the form of inhibition (Steinbeis, 2016, 2018; Steinbeis et al., 2012; Steinbeis & Over, 2017). In addition, decreased temporal discounting has also been linked to cognitive control (Figner et al., 2010; Steinbeis et al., 2016). Thus, these individual non-age-related differences may be explainable by individual differences in EFs. However, I found limited significant relations between EFs and decision-making in the current work.

In Chapter 4, to assess these relations efficiently and reduce the amount of variance and measures, I used latent factors of EFs, which captured the three main domains: inhibition, cognitive flexibility, and working memory (see 4.4.2 Relationships between executive functions and decision-making measures). For inhibition, even though there were initial relationships with pro-social decision-making and delay valuation, these relationships did not survive correction for age. There was an initial relationship between strategic social decision-making and cognitive flexibility, but this relationship did not survive correction. However, there was a significant relationship with the percentage of delayed choices and delay valuation, and both these relationships survived controlling for age. There were initial relationships among pro-social decision-making, delay valuation, and working memory, but these were no longer significant after controlling for age. Thus, in the current work, there do not seem to be as strong relationships between EFs and decision-making as previously expected.

One limitation of the current experimental design for the social and intertemporal decision-making tasks is that while comprehension checks were used before the start of the task, there were no checks afterward. While a common paradigm for these decision-making tasks was followed, I observed that some children during the intertemporal decision-making task either discounted very steeply or not at all, and on occasion, did not follow the expected discounting pattern. Without having quizzed afterward, the reason for this behavior is unclear. It could indicate that these children did not value the rewards offered or the task or believed in the paradigm. Alternatively, it could mean that these children were not sensitive to the delay manipulation. For future work, assessing the motivation for certain decisions afterward would help to understand unexpected patterns of decision-making.

The absence of these strong relations may be due to several things. First, the lack could be since I used latent factors that reduced the variance of the EFs. However, I also conducted a correlation-based analysis between two decision-making factors and the separate EF measures, where I did not see strong relationships between the two (see 4.6.5.6 Decision-making factors and executive functions). Thus, this does not seem solely due to using the latent factors for EFs. Second, it could be due to differences in the task designs for either the EF or the decision-making tasks. However, in the current work, similar task paradigms were used as in a previous study that did find significant relationships between cognitive control and decision-making, such as the Dictator and Ultimatum Game and the Stop Signal Delay Task (see 3.3.3 Cognitive task battery) (Steinbeis et al., 2012, 2016; Steinbeis & Over, 2017). As the cognitive tasks used in the current thesis have previously been reported to have acceptable levels of test-retest reliability, this would suggest that the findings relating to EFs should hold across different studies. However, the test-retest reliability of model-free and model-based decision-making and performance on the two-step task is yet unknown. Future work would be advised to allow assessing the reliability of individual differences in decision-making strategies, to allow reproducible comparisons between studies. Thus, the current thesis shows that EFs may not support pro-social and intertemporal decision-making in 6-13-year-old children. This is mirrored in the null-finding for training-related effects of an inhibitory control paradigm, which will be discussed in the next section.

5.8 Support for the null: interpreting the findings of the cognitive control training paradigm on social and intertemporal decision-making

In Chapter 4, I researched whether social and intertemporal decision-making could be affected by participating in an inhibitory control training paradigm. Training studies have generally been criticized for lack of effectiveness and transferability (Sala & Gobet, 2019). In addition, non-significant results in training studies can mean that there is evidence for the null hypothesis (training does not affect the measure of interest) or that the data are insensitive to providing evidence supporting the null hypothesis (Dienes, 2014; Dougherty et al., 2016).

It is, therefore, of interest in training studies to consider evidence in favor of the null hypothesis. I used the Bayes factor in support of the null hypothesis (cognitive control training did not lead to changes in the decision-making measures) to assess training-related transfer to decision-making tasks. In this case, the Bayes Factor B indicates that the data are B times more likely under the null model than the training model. Therefore, the Bayes Factor allows three different types of conclusions regarding the potential training effects; (i) strong evidence for a null effect (no-training related changes) (B much greater than 1), (ii) strong evidence for the training model (B close to 0); (iii) and the evidence is insensitive (B close to 1) (Dienes, 2014). In Chapter 4, I assess potential training-related changes across three time points split into short-term and long-term effects. The short-term effects focused on potential training-related changes between the pre-training and testing time points immediately after completing the 8-week training paradigm. The long-term effects focused on possible changes between the pre-training and follow-up time points one year after completing the training paradigm.

For most of the measures in the short-term training-related analysis, I found strong evidence supporting the null over the training model. However, for two measures, strategic social decision-making and reward valuation, the Bayes factor did not provide strong evidence for either the null or the training model. Thus, while some ambiguity remains regarding strategic decision-making and reward valuation, I can conclude from these findings that an inhibitory control training paradigm did not lead to changes in pro-social and intertemporal decision-making immediately after training. I also investigated potential long-term training-related effects, and here, for all measures except for strategic social decision-making, did the Bayes evidence strongly support the null.

In sum, in Chapter 4, I did not find as strong a relationship between EFs and pro-social and intertemporal decision-making as expected. In addition, an experimental EF manipulation also did not lead to training-related changes. Thus, the results in this thesis did not seem to be in accordance with previous studies establishing strong relationships between EFs and pro-social and intertemporal decision-making, which provided the foundation of the hypothesis that enhancing EFs via inhibitory control training would lead to further changes in decision-making. Thus, an inhibitory control training paradigm may not effectively improve pro-social and intertemporal decision-making. However, other training approaches could be more effective.

5.9 On effective training paradigms for decision-making

Despite several studies employing EF intervention paradigms, training EFs, such as working memory, cognitive flexibility, or inhibition, have rarely translated to real-world effects (Diamond & Ling, 2016; Holmes et al., 2019; Kable et al., 2017; Simons et al., 2016). Therefore, EF training studies have been criticized for lack of effectiveness and poorly designed paradigms (Sala, Gobet, 2019). Specifically, training studies have been criticized for a) poor definitions of the training mechanism, b) poor design of training protocols, as well as c,) being relatively underpowered to detect small effects (Smid et al., 2020).

A core assumption of training studies has been that training mechanisms are fundamentally related to the outcome measures of interest (Noack et al., 2009). For example, working memory capacity (i.e., the maximal amount of information that can be stored in working memory) correlates highly with general intelligence (measured as fluid reasoning) (Jaeggi et al., 2008). Thus, the logical consequence has been to target working memory capacity to increase intelligence. However, as has been argued previously, two correlated variables, such as working memory span and intelligence, do not necessarily covary when one is artificially being inflated via training because training can tap unshared variance between the two constructs (Moreau & Conway, 2014). Moreover, although relationships between working memory and intelligence might exist at a latent factor level, this is not necessarily the case at the level of single tasks typically used in training studies. EFs are higher-order constructs that include different processes; for instance, working memory consists of storage, rehearsal, and matching, as well as manipulation of information and processing skill. Thus, correlating two tasks does not offer sufficient granularity or direction to identify the true underlying process-based nature of the relationship. Finally, considering task manipulations (e.g., increasing working memory span) as tantamount to training for outcome variables is a nontrivial endeavor. For example, it might not be working memory span that is related to intelligence (Unsworth & Engle, 2005) but rather a shared executive attention-control mechanism required for the active maintenance of information in the face of concurrent processing and inference. Increasing the working memory span may, therefore, not lead to improvements in intelligence (Sala & Gobet, 2017).

In the current work, the training design sought to overcome this by creating a construct of working memory ability based on different working memory tasks (Corsi, n-back tasks) to relate outcome measures to, recruiting a large sample with multiple testing points, and in addition, the training comprised of multiple tasks which used different motor responses to train inhibition (see 4.6.1 Cognitive control training protocol) to obtain the best way to manipulate cognitive control. Furthermore, as training performance is often accompanied by substantial inter and intra-individual differences, the current training design was adaptive to a participant's performance across the different tasks. Lastly, to interpret the evidence for null or training effects properly, I included Bayesian evidence in support of the null to determine whether the training was ineffective or the data insensitive and found that this supported the null. Thus, although the current study used a high-standard training design, was adaptive, and had a considerably large sample (N = 205), I still did not observe training-related effects, and Bayesian evidence strongly supported the null model (no training-related changes).

There is a discussion regarding preference for domain-specific or domain-general training studies. It has been argued elsewhere that training paradigms with the goal of far-transfer would be more effective when embedded into the domain they would like to influence (P. Peng & Lee Swanson, 2022; Ramani et al., 2017). For example, an intervention study centering around inhibitory control aimed to improve counterintuitive reasoning in math and science used a paradigm where the training was embedded into the target groups' science and math curriculum and observed modest training-related changes in science achievement (Wilkinson et al., 2019). Thus, if the goal is to influence pro-social decision-making, training paradigms that focus on, for example, increasing empathy may have more potential to lead to training-related changes. Alternatively, creating a training paradigm with improved ecological validity, for example, real-world social interactions, may lead to increases in pro-social decision-making (de Felice et al., 2021), which could potentially be aided by virtual reality environments (Pan & Hamilton, 2018). A previous study in childhood that used behavioral control in the context of a social story-telling setting did observe increases in pro-social decision-making (Steinbeis & Over, 2017). Thus, moving to a more ecologically valid paradigm embedded in the decision-making realm of interest may achieve transfer effects in decision-making.

Alternatively, one-shot economic games such as the Dictator and Ultimatum Game may not be the best outcome measure to attempt and influence via an EF training paradigm. In one-shot economic games, participants only provide a single measure of decision-making. In contrast, decision-making tasks with more trials may provide more insight into the underlying social and intertemporal preferences (Lockwood & Klein-Flügge, 2021). In particular, reinforcement learning tasks that combine a social component would be able to investigate whether pro-social learning could be affected by an EF training paradigm (Lockwood et al., 2016; Pauli et al., 2022). Interestingly, previous work has shown that, especially when we are young, we may learn faster when we make decisions differently that affect ourselves rather than someone else (Cutler et al., 2021; Lockwood et al., 2016, 2021). In addition, other studies have suggested that especially inhibitory control may be needed to suppress the prepotent impulse to benefit ourselves (Steinbeis, 2018; Steinbeis et al., 2012; Steinbeis & Over, 2017). For example, as in Chapter 4, increased reaction times were linked to more pro-social decisions and less steep temporal discounting; this could indicate that the ability to withhold prepotent responses and deliberate over an action may lead to desired improvements in decision-making. Thus, decision-making tasks that involve social learning may be a better candidate to test whether EF training also leads to changes in social decision-making.

In sum, in the current work, a randomized-controlled domain-general and adaptive inhibitory control training paradigm did not lead to training-related changes in pro-social and intertemporal decision-making. While the current training was well designed and sought to counter the common shortcomings of training paradigms, other training approaches could be more effective in translating to changes in decision-making. Future studies should aim to take over the benefits of the current training paradigm in (i) having a range of different tasks that tackle the same mechanisms (in this case, inhibitory control), (ii) being adaptive to individual performance, and (iii) having a similarly large sample size. However, future avenues for training could improve by exploring a) more ecologically valid training paradigms embedded in the relevant realm of decision-making and b) using decision-making outcome measures that offer a better assessment of potential training-related changes, such as changes in reaction time.

5.10 Conclusion

This thesis sought to investigate value-based decision-making in childhood from the perspective of reinforcement learning and social and intertemporal decision-making. In the first Chapter, I discussed the three dominant theoretical branches of decision-making research: normative, descriptive, and prescriptive theories. Next, I discussed how research into decision-making moved from assuming people are rational decision-makers to accepting that real people make irrational decisions. Finally, I introduced the concept of bounded rationality and how constraints on the number of alternatives we can consider or how many contexts can be reflected on are thought to bind the rationality in our decisions. The constraints on our rationality can be interpreted as executive functions (EFs); our working memory constrains how much information we can manipulate in our minds, whereas cognitive control allows us to select the right option every time, and cognitive flexibility can enable us to switch between options flexibly. To this end, I investigate reinforcement learning and social and intertemporal decision-making from the context of executive functions to investigate whether these cognitive abilities may support rational decision-making. In sum, I found that the relationships between EFs and decision-making were not as robust as expected.

6. 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

Allgaier, K., Ścigała, K. A., Trautwein, U., Hilbig, B. E., & Zettler, I. (2020). Honesty-humility and dictator and ultimatum game-giving in children. *Journal of Research in Personality*, *85*. https://doi.org/10.1016/j.jrp.2019.103907

Alloway, T. P., & Alloway, R. G. (2009). The efficacy of working memory training in improving crystallized intelligence. *Nature Precedings*, *January 2009*. https://doi.org/10.1038/npre.2009.3697.1

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, *17*(7), 1493–1503. https://doi.org/10.1093/cercor/bhl078

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69. https://doi.org/10.1038/npp.2009.131

Basile, A. G., & Toplak, M. E. (2015). Four converging measures of temporal discounting and their relationships with intelligence, executive functions, thinking dispositions, and behavioral outcomes. *Frontiers in Psychology*, *6*(JUN), 1–13. https://doi.org/10.3389/fpsyg.2015.00728

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

Bauer, M., Chytilová, J., & Pertold-Gebicka, B. (2014). Parental background and other-regarding preferences in children. *Experimental Economics*, *17*(1), 24–46. https://doi.org/10.1007/s10683-013-9355-y

Beierholm, U. R., Anen, C., Quartz, S., & Bossaerts, P. (2011). Separate encoding of model-based and model-free valuations in the human brain. *NeuroImage*, *58*(3), 955–962. https://doi.org/10.1016/j.neuroimage.2011.06.071

Bellman, R. (1954). Dynamic programming and a new formalism in the calculus of variations. *Proceedings of the National Academy of Sciences*, *40*(4), 231–235.

Bellman, R., & Dreyfus, S. E. (1959). Functional approximations and dynamic programming. *Math Tables and Other Aides to Computation*, *13*, 247–251.

Bellman, R. E. (1957a). A Markov decision process. *Journal of Mathematical Mech*, *6*, 679–684.

Bellman, R. E. (1957b). *Dynamic Programming*. Princeton University Press.

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior*, *28*(3), 168–175. https://doi.org/10.1016/j.evolhumbehav.2006.10.003

Bengtsson, H. (2021). A Unifying Framework for Parallel and Distributed Processing in R using Futures. *R Journal*, *13*(2), 273–291. https://doi.org/10.32614/RJ-2021-048

Ben-Shachar, M. S., Makowski, D., & Lüdecke, 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

Bentler, P. (1990). Comparitive fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238.

Bernhardt, B. C., Klimecki, O. M., Leiberg, S., & Singer, T. (2014). Structural covariance networks of the dorsal anterior insula predict females' individual differences in empathic responding. *Cerebral Cortex*, *24*(8), 2189–2198. https://doi.org/10.1093/cercor/bht072

Bernhardt, B. C., Smallwood, J., Tusche, A., Ruby, F. J. M., Engen, H. G., Steinbeis, N., & Singer, T. (2014). Medial prefrontal and anterior cingulate cortical thickness predicts shared individual differences in self-generated thought and temporal discounting. *NeuroImage*, *90*, 290–297. https://doi.org/10.1016/j.neuroimage.2013.12.040

Bickel, W. K., Koffarnus, M. N., Moody, L., & Wilson, A. G. (2014). The behavioral- and neuro-economic process of temporal discounting: A candidate behavioral marker of addiction. *Neuropharmacology*, *76*(PART B), 518–527. https://doi.org/10.1016/j.neuropharm.2013.06.013

Bickel, W. K., Miller, M. L., Yi, R., Kowal, B. P., Lindquist, D. M., & Pitcock, J. A. (2007). Behavioral and neuroeconomics of drug addiction: Competing neural systems and temporal discounting processes. *Drug and Alcohol Dependence*, *90*(SUPPL. 1), 85–91. https://doi.org/10.1016/j.drugalcdep.2006.09.016

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Blake, P. R., & McAuliffe, K. (2011). “I had so much it didn’t seem fair” Eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215–224. https://doi.org/10.1016/j.cognition.2011.04.006

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge-behavior gap. *Trends in Cognitive Sciences*, *18*(11), 559–561. https://doi.org/10.1016/j.tics.2014.08.003

Böckler, A., Tusche, A., & Singer, T. (2016). The Structure of Human Prosociality: Differentiating Altruistically Motivated, Norm Motivated, Strategically Motivated, and Self-Reported Prosocial Behavior. *Social Psychological and Personality Science*, *7*(6), 530–541. https://doi.org/10.1177/1948550616639650

Bolenz, F., & Eppinger, B. (2021). Valence bias in metacontrol of decision making in adolescents and young adults. *Child Development*, 1–14. https://doi.org/10.1111/cdev.13693

Bolenz, F., Kool, W., Reiter, A., & Eppinger, B. (2019). Metacontrol of decision-making strategies in human aging. *ELife*, *8*. https://doi.org/10.7554/eLife.49154

Bos, W. van den, & McClure, S. M. (2013). Towards a general model of temporal discounting. *Journal of the Experimental Analysis of Behavior*, *99*(1), 58–73. https://doi.org/10.1002/jeab.6

Bourdieu, P. (1985). The forms of capital. In J. G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood.

Boureau, Y. L., Sokol-Hessner, P., & Daw, N. D. (2015). Deciding How To Decide: Self-Control and Meta-Decision Making. *Trends in Cognitive Sciences*, *19*(11), 700–710. https://doi.org/10.1016/j.tics.2015.08.013

Brosnan, S. F., Houser, D., Leimgruber, K., Xiao, E., Chen, T., & de Waal, F. B. M. (2010). Competing demands of prosociality and equity in monkeys. *Evolution and Human Behavior*, *31*(4), 279–288. https://doi.org/10.1016/j.evolhumbehav.2010.02.003

Bull, J. J., & Rice, W. R. (1991). Distinguishing mechanisms for the evolution of co-operation. *Journal of Theoretical Biology*, *149*(1), 63–74. https://doi.org/10.1016/S0022-5193(05)80072-4

Bulley, A., Henry, J., & Suddendorf, T. (2016). Prospection and the present moment: The role of episodic foresight in intertemporal choices between immediate and delayed rewards. *Review of General Psychology*, *20*(1), 29–47. https://doi.org/10.1037/gpr0000061

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Burnham, J. C. (1972). Thorndike's puzzle boxes. *Journal of the History of the Behavioral Sciences*.

Buss, A. T., & Spencer, J. P. (2014). The emergent executive: a dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, *79*(2). https://doi.org/10.1002/mono.12096

Buss, A. T., & Spencer, J. P. (2018). Changes in frontal and posterior cortical activity underlie the early emergence of executive function. *Developmental Science*, *21*(4), 1–14. https://doi.org/10.1111/desc.12602

Calderon, J., & Newburger, J. W. (2018). Working Memory Training. *Critical Care Medicine*, *46*(7), 1199–1201. https://doi.org/10.1097/ccm.0000000000003172

Chajes, J. R., Grossmann, T., & Vaish, A. (2022). Fairness takes time: Development of cooperative decision making in fairness context. *Journal of Experimental Child Psychology*, *216*, 105344. https://doi.org/10.1016/j.jecp.2021.105344

Chandler, J. (2017). Descriptive decision theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Lab, Stanford University.

Chapman, G. B. (1996). Temporal Discounting and Utility for Health and Money. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 22, Issue 3).

Chen, Y. N., Mitra, S., & Schlaghecken, F. (2008). Sub-processes of working memory in the N-back task: An investigation using ERPs. *Clinical Neurophysiology*, *119*(7), 1546–1559. https://doi.org/10.1016/j.clinph.2008.03.003

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

Chevalier, N. (2018). Willing to Think Hard? The Subjective Value of Cognitive Effort in Children. *Child Development*, *89*(4), 1283–1295. https://doi.org/10.1111/cdev.12805

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool Executive Functioning Abilities Predict Early Mathematics Achievement. *Developmental Psychology*, *46*(5), 1176–1191. https://doi.org/10.1037/a0019672

Collins, A. G. E., & Cockburn, J. (2020). Beyond dichotomies in reinforcement learning. *Nature Reviews Neuroscience*, *21*(10), 576–586. https://doi.org/10.1038/s41583-020-0355-6

Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The role of psychometrics in individual differences research in cognition: A case study of the AX-CPT. *Frontiers in Psychology*, *8*(SEP), 1–16. https://doi.org/10.3389/fpsyg.2017.01482

Coutureau, E., & di Scala, G. (2009). Entorhinal cortex and cognition. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *33*(5), 753–761. https://doi.org/10.1016/j.pnpbp.2009.03.038

Cremer, A., Kalbe, F., Gläscher, J., & Schwabe, L. (2021). Stress reduces both model-based and model-free neural computations during flexible learning. *NeuroImage*, *229*(January), 117747. https://doi.org/10.1016/j.neuroimage.2021.117747

Cutler, J., Wittmann, M. K., Abdurahman, A., Hargitai, L. D., Drew, D., Husain, M., & Lockwood, P. L. (2021). Ageing is associated with disrupted reinforcement learning whilst learning to help others is preserved. *Nature Communications*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-24576-w

Dalley, J. W., Everitt, B. J., & Robbins, T. W. (2011). Impulsivity, Compulsivity, and Top-Down Cognitive Control. *Neuron*, *69*(4), 680–694. https://doi.org/10.1016/j.neuron.2011.01.020

Daugherty, J. R., & Brase, G. L. (2010). Taking time to be healthy: Predicting health behaviors with delay discounting and time perspective. *Personality and Individual Differences*, *48*(2), 202–207. https://doi.org/10.1016/j.paid.2009.10.007

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

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Daw, N. D. (2018). Are we of two minds? *Nature Neuroscience*, *21*(November), 1497–1499. https://doi.org/10.1038/s41593-018-0258-2

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

Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, *446*(7137), 794–796. https://doi.org/10.1038/nature05651

de Felice, S., Vigliocco, G., & Hamilton, A. F. de C. (2021). Social interaction is a catalyst for adult human learning in online contexts. *Current Biology*, *31*(21), 4853-4859.e3. https://doi.org/10.1016/j.cub.2021.08.045

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

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Dezfouli, A., & Balleine, B. W. (2013). Actions, Action Sequences and Habits: Evidence That Goal-Directed and Habitual Action Control Are Hierarchically Organized. *PLoS Computational Biology*, *9*(12). https://doi.org/10.1371/journal.pcbi.1003364

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1002/9781119557036.ch10

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old (Science (959)). *Science*, *334*(6054), 311. https://doi.org/10.1126/science.334.6054.311-d

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, *18*, 34–48. https://doi.org/10.1016/j.dcn.2015.11.005

Dickinson, A. (1985). Actions and habits : the development of behavioural autonomy. *Biological Sciences*, *78*(308(1135)), 67–78.

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

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*(July), 1–17. https://doi.org/10.3389/fpsyg.2014.00781

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

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences*, *1*, 101–106. https://doi.org/10.1016/j.cobeha.2014.10.007

Donati, M. A., Panno, A., Chiesi, F., & Primi, C. (2014). A mediation model to explain decision making under conditions of risk among adolescents: The role of fluid intelligence and probabilistic reasoning. *Journal of Clinical and Experimental Neuropsychology*, *36*(6), 588–595. https://doi.org/10.1080/13803395.2014.918091

Dougherty, M. R., Hamovitz, T., & Tidwell, J. W. (2016). Reevaluating the effectiveness of n-back training on transfer through the Bayesian lens: Support for the null. *Psychonomic Bulletin and Review*, *23*(1), 306–316. https://doi.org/10.3758/s13423-015-0865-9

Dow, S. C. (1990). Beyond dualism. In *Cambridge Journal of Economics* (Vol. 14, pp. 143–157). Oxford University Press.

Druzgal, T. J., & D'Esposito, M. (2001). Activity in fusiform face area modulated as a function of working memory load. *Cognitive Brain Research*, *10*(3), 355–364. https://doi.org/10.1016/S0926-6410(00)00056-2

Dubois, M., Bowler, A., Moses-Payne, M. E., Habicht, J., Moran, R., Steinbeis, N., & Hauser, T. U. (2022). Exploration heuristics decrease during youth. *Cognitive, Affective, & Behavioral Neuroscience*, *0123456789*. https://doi.org/10.3758/s13415-022-01009-9

Edele, A., Dziobek, I., & Keller, M. (2013). Explaining altruistic sharing in the dictator game: The role of affective empathy, cognitive empathy, and justice sensitivity. *Learning and Individual Differences*, *24*, 96–102. https://doi.org/10.1016/j.lindif.2012.12.020

Elliott, E. (2019). Normative Decision Theory. *Analysis*, *79*(4), 755–772. https://doi.org/10.1093/analys/anz059

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

Erez, Y., & Reyna, V. F. (2019). Decision-Making. In R. J. Sternberg & W. E. Pickren (Eds.), *The Cambridge Handbook of the Intellectual History of Psychology* (pp. 216–249). Cambridge University Press. https://doi.org/10.1017/9781108290876.009

Farrell Pagulayan, K., Busch, R., Medina, K., Bartok, J., & Krikorian, R. (2006). Developmental normative data for the Corsi Block-Tapping task. *Journal of Clinical and Experimental Neuropsychology*, *28*(6), 1043–1052. https://doi.org/10.1080/13803390500350977

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.

Feher, C., & Hare, T. A. (2019). *Model-free or muddled models in the two-stage task ?* 1–26.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079–1083. https://doi.org/10.1038/nature07155

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791. https://doi.org/10.1038/nature02043

Fehr, E., & Gachter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *American Economic Association*, *90*(4), 980–994.

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., & Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, *13*(5), 538–539. https://doi.org/10.1038/nn.2516

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X

Fiske, A., & Holmboe, K. (2019). Neural substrates of early executive function development. *Developmental Review*, *52*(May), 42–62. https://doi.org/10.1016/j.dr.2019.100866

Flora, D. B. (2020). Your Coefficient Alpha Is Probably Wrong, but Which Coefficient Omega Is Right? A Tutorial on Using R to Obtain Better Reliability Estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484–501. https://doi.org/10.1177/2515245920951747

Ganesan, K., & Steinbeis, N. (2021). Effort-Related Decision-Making And Its Underlying Processes During Childhood. *Developmental Psychology*, *57*(9), 1487–1496. https://doi.org/10.1037/dev0001228

Ganesan, K., & Steinbeis, N. (2022). Development and plasticity of executive functions: A value-based account. *Current Opinion in Psychology*, *44*, 215–219. https://doi.org/10.1016/j.copsyc.2021.09.012

Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychology*, *20*(6), 713–736. https://doi.org/10.1080/09297049.2013.857650

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

Gershman, S. J. (2018). *"mfit": simple model-fitting tools*. https://github.com/sjgershm/mfit

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective and Behavioral Neuroscience*, *15*(3), 523–536. https://doi.org/10.3758/s13415-015-0347-6

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

Gobet, F., & Sala, G. (2022). Cognitive training: a field in search of a phenomenon. *Perspectives on Psychological Science*. http://eprints.lse.ac.uk/113888/

Gonthier, C., Macnamara, B. N., Chow, M., Conway, A. R. A., & Braver, T. S. (2016). Inducing proactive control shifts in the AX-CPT. *Frontiers in Psychology*, *7*(NOV), 1–14. https://doi.org/10.3389/fpsyg.2016.01822

Green, L., Myerson, J., & Ostaszewski, P. (1999). Discounting of delayed rewards across the life span. *Behavioural Processes*, *46*, 89–96.

Guillou, L., Grandin, A., & Chevallier, C. (2020). *Correcting misperceptions of relative income: impact on temporal discounting and social trust.* 1–28.

Guillou, L., Grandin, A., & Chevallier, C. (2021). Temporal discounting mediates the relationship between socio-economic status and social trust. *Royal Society Open Science*, *8*(6). https://doi.org/10.1098/rsos.202104

Gummerum, M., Hanoch, Y., & Keller, M. (2008). When child development meets economic game theory: An interdisciplinary approach to investigating social development. *Human Development*, *51*(4), 235–261. https://doi.org/10.1159/000151494

Gummerum, M., Hanoch, Y., Keller, M., Parsons, K., & Hummel, A. (2010). Preschoolers' allocations in the dictator game: The role of moral emotions. *Journal of Economic Psychology*, *31*(1), 25–34. https://doi.org/10.1016/j.joep.2009.09.002

Guo, Q., & Feng, L. (2017). The associations between perceived parenting styles, empathy, and altruistic choices in economic games: A study of Chinese children. *Frontiers in Psychology*, *8*(OCT), 1–11. https://doi.org/10.3389/fpsyg.2017.01843

Guo, R., & Wu, Z. (2021). Empathy as a buffer: How empathy moderates the emotional effects on Preschoolers' sharing. *British Journal of Psychology*, *112*(2), 412–432. https://doi.org/10.1111/bjop.12466

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, *3*(4), 367–388.

Habicht, J., Bowler, A., Moses-Payne, M., & Hauser, T. U. (2022). 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*, *151*(8), 1843–1853. https://doi.org/10.1037/xge0001138

Hamilton, K. R., & Potenza, M. N. (2012). Relations among delay discounting, addictions, and money mismanagement: Implications and future directions. *American Journal of Drug and Alcohol Abuse*, *38*(1), 30–42. https://doi.org/10.3109/00952990.2011.643978

Heckman, J. J. (2006). Investing in Disadvantaged Children. *Social Sciences*, *312*(June), 2005–2007.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q., & Tracer, D. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in

15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795–815. https://doi.org/10.1017/S0140525X05000142

Herne, K., Hietanen, J. K., Lappalainen, O., & Palosaari, E. (2022). The influence of role awareness, empathy induction and trait empathy on dictator game giving. *PLoS ONE*, *17*(3 March), 1–19. https://doi.org/10.1371/journal.pone.0262196

Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S. A., & Zettler, I. (2015). From personality to altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of Research in Personality*, *55*, 46–50. https://doi.org/10.1016/j.jrp.2014.12.004

Holmes, J., Woolgar, F., Hampshire, A., & Gathercole, S. E. (2019). Are working memory training effects paradigm-specific? *Frontiers in Psychology*, *10*(MAY), 1–10. https://doi.org/10.3389/fpsyg.2019.01103

Hughes, C., & Ensor, R. (2010). Do early social cognition and executive function predict individual differences in preschoolers' prosocial and antisocial behavior? In *Self and Social Regulation: Social interaction and the development of social understanding and executive functions*. Oxford University Press.

Insel, C., Glenn, C. R., Nock, M. K., & Sommerville, L. H. (2019). Aberrant Striatal Tracking of Reward Magnitude in Youth With Current or Past-Year Depression. *Nature Communications*, *8*(1), 44. https://doi.org/10.1523/JNEUROSCI.1010-06.2006

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). 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

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jeffreys, H. (1961). *The Theory of Probability* (1st/3rd Ed). Oxford University Press.

Jessen, F., Feyen, L., Freymann, K., Tepest, R., Maier, W., Heun, R., Schild, H. H., & Scheef, L. (2006). Volume reduction of the entorhinal cortex in subjective memory impairment. *Neurobiology of Aging*, *27*(12), 1751–1756. https://doi.org/10.1016/j.neurobiolaging.2005.10.010

Johann, V. E., & Karbach, J. (2020). Effects of game-based and standard executive control training on cognitive and academic abilities in elementary school children. *Developmental Science*, *23*(4), 1–18. https://doi.org/10.1111/desc.12866

Jones, A., di Lemma, L. C. G., Robinson, E., Christiansen, P., Nolan, S., Tudur-Smith, C., & Field, M. (2016). Inhibitory control training for appetitive behaviour change: A meta-analytic investigation of mechanisms of action and moderators of effectiveness. *Appetite*, *97*, 16–28. https://doi.org/10.1016/j.appet.2015.11.013

Judd, N., & Klingberg, T. (2021). Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour*, *5*(11), 1548–1554. https://doi.org/10.1038/s41562-021-01118-4

Kable, J. W., Caulfield, M. K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., Cooper, N., Ashare, R., Audrain-McGovern, J., Hornik, R., Diefenbach, P., Lee, F. J., & Lerman, C. (2017). No effect of commercial cognitive training on brain activity, choice behavior, or cognitive performance. *Journal of Neuroscience*, *37*(31), 7390–7402. https://doi.org/10.1523/JNEUROSCI.2832-16.2017

Kacelnik, A. (1997). *Normative and descriptive models of decision making: time discounting and risk sensitivity*.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, *93*(5), 1449–1475.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*(6), 978–990. https://doi.org/10.1111/j.1467-7687.2009.00846.x

Karbach, J., Strobach, T., & Schubert, T. (2015). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychology*, *21*(3), 285–301. https://doi.org/10.1080/09297049.2014.899336

Karbach, J., & Verhaeghen, P. (2014). Making Working Memory Work: A Meta-Analysis of Executive-Control and Working Memory Training in Older Adults. *Psychological Science*, *25*(11), 2027–2037. https://doi.org/10.1177/0956797614548725

Kassai, R., Futo, J., Demetrovics, Z., & Takacs, Z. K. (2019). A meta-analysis of the experimental evidence on the near- and far-transfer effects among children's executive function skills. *Psychological Bulletin*, *145*(2), 165–188. https://doi.org/10.1037/bul0000180

Keidel, K., Rramani, Q., Weber, B., Murawski, C., & Ettinger, U. (2021). Individual Differences in Intertemporal Choice. *Frontiers in Psychology*, *12*(April). https://doi.org/10.3389/fpsyg.2021.643670

Kenny, L., Hill, E., & Hamilton, A. F. de C. (2016). The Relationship between Social and Motor Cognition in Primary School Age-Children. *Frontiers in Psychology*, *7*(February), 1–12. https://doi.org/10.3389/fpsyg.2016.00228

Keramati, M., Dezfouli, A., & Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Computational Biology*, *7*(5). https://doi.org/10.1371/journal.pcbi.1002055

Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(45), 12868–12873. https://doi.org/10.1073/pnas.1609094113

Kim, H., Ku, B., Kim, J. Y., Park, Y. J., & Park, Y. B. (2016). Confirmatory and exploratory factor analysis for validating the phlegm pattern questionnaire for healthy subjects. *Evidence-Based Complementary and Alternative Medicine*, *2016*. https://doi.org/10.1155/2016/2696019

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

Kirby, K. N. (2009). One-year temporal stability of delay-discount rates. *Psychonomic Bulletin and Review*, *16*(3), 457–462. https://doi.org/10.3758/PBR.16.3.457

Klingberg, T. (2005). Computerized training of working memory in children with ADHD. *J. Am. Acad. Child Adolesc. Psychiatry*, *44*(2), 177–186. https://doi.org/10.1016/s0924-977x(07)70216-x

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

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829–832. https://doi.org/10.1126/science.1129156

Koenigs, M., Barbey, A. K., Postle, B. R., & Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *Journal of Neuroscience*, *29*(47), 14980–14986. https://doi.org/10.1523/JNEUROSCI.3706-09.2009

Könen, T., & Karbach, J. (2015). The benefits of looking at intraindividual dynamics in cognitive training data. *Frontiers in Psychology*, *6*(MAY), 4–7. https://doi.org/10.3389/fpsyg.2015.00615

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/a0031048

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

Kool, W., Gershman, S. J., & Cushman, F. A. (2018). Planning Complexity Registers as a Cost in Metacontrol. *Journal of Cognitive Neuroscience*, *30*(10), 1391–1404. https://doi.org/10.1162/jocn

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*(May), 1–9. https://doi.org/10.1038/ncomms8455

Lee, J. J., & Keramati, M. (2017). Flexibility to contingency changes distinguishes habitual and goal-directed strategies in humans. *PLoS Computational Biology*, *13*(9), 1–15. https://doi.org/10.1371/journal.pcbi.1005753

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, *81*(3), 687–699. https://doi.org/10.1016/j.neuron.2013.11.028

Lengersdorff, L. L., Wagner, I. C., Lockwood, P. L., & Lamm, C. (2020). When implicit prosociality trumps selfishness: The neural valuation system underpins more optimal choices when learning to avoid harm to others than to oneself. *Journal of Neuroscience*, *40*(38), 7286–7299. https://doi.org/10.1523/JNEUROSCI.0842-20.2020

Lerch, J. P., & Evans, A. C. (2005). Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage*, *24*(1), 163–173. https://doi.org/10.1016/j.neuroimage.2004.07.045

Li, L., Rizzo, M. T., Burkholder, A. R., & Killen, M. (2017). Theory of mind and resource allocation in the context of hidden inequality. *Cognitive Development*, *43*, 25–36. https://doi.org/10.1016/j.cogdev.2017.02.001

Li, S. (2003). The role of expected value illustrated in decision-making under risk: Single-play vs multiple-play. *Journal of Risk Research*, *6*(2), 113–124. https://doi.org/10.1080/1366987032000078893

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47–56. https://doi.org/10.1016/j.jneumeth.2016.03.001

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 1–85. https://doi.org/10.1017/S0140525X1900061X

Lockwood, P. L., Apps, M. A. J., Valton, V., Viding, E., & Roiser, J. P. (2016). Neurocomputational mechanisms of prosocial learning and links to Empathy. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(35), 9763–9768. https://doi.org/10.1073/pnas.1603198113

Lockwood, P. L., Hamonet, M., Zhang, S. H., Ratnavel, A., Salmony, F. U., Husain, M., & Apps, M. A. J. (2017). Prosocial apathy for helping others when effort is required. *Nature Human Behaviour*, *1*(7), 1–10. https://doi.org/10.1038/s41562-017-0131

Lockwood, P. L., & Klein-Flügge, M. C. (2021). Computational modelling of social cognition and behaviour - A reinforcement learning primer. *Social Cognitive and Affective Neuroscience*, *16*(8), 761–771. https://doi.org/10.1093/scan/nsaa040

Lockwood, P. L., Klein-Flügge, M. C., Abdurahman, A., & Crockett, M. J. (2020). Model-free decision making is prioritized when learning to avoid harming others. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(44), 27719–27730. https://doi.org/10.1073/pnas.2010890117

Lockwood, P. L., & Wittmann, M. K. (2018). Ventral anterior cingulate cortex and social decision-making. *Neuroscience and Biobehavioral Reviews*, *92*, 187–191. https://doi.org/10.1016/j.neubiorev.2018.05.030

Lockwood, P. L., Wittmann, M., Nili, H., Matsumoto-Ryan, M., Abdurahman, A., Cutler, J., Husain, M., & Apps, M. A. J. (2021). Distinct neural representations for prosocial and self-benefitting effort. *BioRxiv*, 2021.09.27.461936. https://www.biorxiv.org/content/10.1101/2021.09.27.461936v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.09.27.461936v1.abstract

Loose, L. S., Wisniewski, D., Rusconi, M., Goschke, T., & Haynes, J. D. (2017). Switch-independent task representations in frontal and parietal cortex. *Journal of Neuroscience*, *37*(33), 8033–8042. https://doi.org/10.1523/JNEUROSCI.3656-16.2017

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

Lucas, M. M., Wagner, L., & Chow, C. (2008). Fair game: The intuitive economics of resource exchange in four-year olds. *Journal of Social, Evolutionary, and Cultural Psychology*, *2*(3), 74–88. https://doi.org/10.1037/h0099353

Lüdecke, 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

Luna, B., & Sweeney, J. A. (2004). The emergence of collaborative brain function: fMRI studies of the development of response inhibition. *Annals of the New York Academy of Sciences*, *1021*(1), 296–309.

Luna, B., Thulborn, K. R., Munoz, D. P., Merriam, E. P., Garver, K. E., Minshew, N. J., Keshavan, M. S., Genovese, C. R., Eddy, W. F., & Sweeney, J. A. (2001). Maturation of widely distributed brain function subserves cognitive development. *NeuroImage*, *13*(5), 786–793. https://doi.org/10.1006/nimg.2000.0743

Machina, M. J. (1987). Decision-making in the presence of risk. *Science*, *236*(4801), 537–543.

Marc, J. M. (2020). *AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c)* (R package version 2.3-1). https://cran.r-project.org/package=AICcmodavg

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., … Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), 654–660. https://doi.org/10.1038/s41586-022-04492-9

Matsumoto, Y., Yamagishi, T., Li, Y., & Kiyonari, T. (2016). Prosocial behavior increases with age across five economic games. *PLoS ONE*, *11*(7), 1–16. https://doi.org/10.1371/journal.pone.0158671

Matzke, D., Verbruggen, F., & Logan, G. D. (2018). The Stop-Signal Paradigm. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. https://doi.org/10.1002/9781119170174.epcn510

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, *1*(2). https://doi.org/10.1038/s41562-016-0042

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental Science*, *24*(4). https://doi.org/10.1111/desc.13095

Mellis, A. M., Snider, S. E., Deshpande, H. U., LaConte, S. M., & Bickel, W. K. (2019). Practicing prospection promotes patience: Repeated episodic future thinking cumulatively reduces delay discounting. *Drug and Alcohol Dependence*, *204*(May), 107507. https://doi.org/10.1016/j.drugalcdep.2019.06.010

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Miller, K. J., Ludvig, E. A., Pezzulo, G., & Shenhav, A. (2018). Realigning models of habitual and goal-directed decision-making. In *Goal-Directed Decision Making: Computations and Neural Circuits*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-812098-9.00018-8

Minsky, M. (1961). Steps towards artificial intelligence. *Proceedings of the IRE*, *49*(1), 8–30.

Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., Fryer, T. D., Williams, G. B., Hodges, J. R., & Nestor, P. J. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain*, *133*(11), 3256–3268. https://doi.org/10.1093/brain/awq272

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology*, *54*(4), 687.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H. L., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

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

Montez, D. F., Calabro, F. J., & Luna, B. (2017). The expression of established cognitive brain states stabilizes with working memory development. *ELife*, *6*, 1–26. https://doi.org/10.7554/eLife.25606

Moreau, D., & Conway, A. R. A. (2014). The case for an ecological approach to cognitive training. *Trends in Cognitive Sciences*, *18*(7), 334–336. https://doi.org/10.1016/j.tics.2014.03.009

Moutoussis, M., Garzón, B., Neufeld, S., Bach, D. R., Rigoli, F., Goodyer, I., Bullmore, E., Fonagy, P., Jones, P., Hauser, T., Romero-Garcia, R., St Clair, M., Vértes, P., Whitaker, K., Inkster, B., Prabhu, G., Ooi, C., Toseeb, U., Widmer, B., … Dolan, R. J. (2021). Decision-making ability, psychopathology, and brain connectivity. *Neuron*, *109*(12), 2025-2040.e7. https://doi.org/10.1016/j.neuron.2021.04.019

Niebaum, J. C., Chevalier, N., Guild, R. M., & Munakata, Y. (2019). Adaptive control and the avoidance of cognitive control demands across development. *Neuropsychologia*, *123*(February 2018), 152–158. https://doi.org/10.1016/j.neuropsychologia.2018.04.029

Niebaum, Jesse C., Chevalier, N., Guild, R.M., Munakata, Y. (2020). *Developing Adaptive Control: Age-related differences in task choices and awareness of proactive and reactive control demands*.

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. *Restorative Neurology and Neuroscience*, *27*(5), 435–453. https://doi.org/10.3233/RNN-2009-0496

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., Evans, M., & Hartley, C. (2020). *Moving developmental research online: comparing in-lab and web-based studies of model-based reinforcement learning*. *6*, 1–18. https://doi.org/10.31219/osf.io/vewyq

O'Doherty, J. P., Lee, S. W., & McNamee, D. (2015). The structure of reinforcement-learning mechanisms in the human brain. *Current Opinion in Behavioral Sciences*, *1*, 94–100. https://doi.org/10.1016/j.cobeha.2014.10.004

Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The Curse of Planning: Dissecting Multiple Reinforcement-Learning Systems by Taxing the Central Executive. *Psychological Science*, *24*(5), 751–761. https://doi.org/10.1177/0956797612463080

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. (2015). Cognitive Control Predicts Use of Model-based Reinforcement Learning. *Journal of Cognitive Neuroscience*, *27*(2), 319–333.

Paap, K. R., & Oliver, S. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, *274*, 81–93. https://doi.org/10.1016/j.jneumeth.2016.10.002

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

Pan, X., & Hamilton, A. F. de C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

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

Pauli, R., Brazil, I., Kohls, G., Klein-Flügge, M. C., Rogers, J. C., Dikeos, D., Dochnal, R., Fairchild, G., Fernández-Rivas, A., Herpertz-Dahlmann, B., Hervas, A., Konrad, K., Popma, A., Stadler, C., Freitag, C. M., de Brito, S. A., & Lockwood, P. L. (2022). Action initiation and punishment learning differ from childhood to adolescence while reward learning remains stable. *BioRxiv Preprint*, 1–32.

Paulus, M., Nöth, A., & Wörle, M. (2018). Preschoolers' resource allocations align with their normative judgments. *Journal of Experimental Child Psychology*, *175*, 117–126. https://doi.org/10.1016/j.jecp.2018.05.001

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.1289/EHP4713

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peng, P., & Lee Swanson, H. (2022). The domain-specific approach of working memory training. *Developmental Review*, *65*, 101035. https://doi.org/10.1016/j.dr.2022.101035

Peng, X., & Burwell, R. D. (2021). Beyond the hippocampus: The role of parahippocampal-prefrontal communication in context-modulated behavior. *Neurobiology of Learning and Memory*, *185*(June), 107520. https://doi.org/10.1016/j.nlm.2021.107520

Penner, I. K., Kobel, M., Stcklin, M., Weber, P., Opwis, K., & Calabrese, P. (2012). The stroop task: Comparison between the original paradigm and computerized versions in children and adults. *Clinical Neuropsychologist*, *26*(7), 1142–1153. https://doi.org/10.1080/13854046.2012.713513

Peters, J., & Büchel, C. (2011). The neural mechanisms of inter-temporal decision-making: Understanding variability. *Trends in Cognitive Sciences*, *15*(5), 227–239. https://doi.org/10.1016/j.tics.2011.03.002

Plonsky, O., & Erev, I. (2021). To predict human choice, consider the context. *Trends in Cognitive Sciences*, *xx*(xx), 1–2. https://doi.org/10.1016/j.tics.2021.07.007

Portes, A. (1998). SOCIAL CAPITAL: Its Origins and Applications in Modern Sociology. *Annu. Rev. Sociol*, *24*, 1–24. www.annualreviews.org

Posner, M. I., & Snyder, C. R. R. (1975). Attention and Cognitive Control. In R. L. Solso (Ed.), *Information Processing and Cognition* (pp. 55–85). Erlbaum.

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

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental Child Psychology*, *108*(3), 621–637. https://doi.org/10.1016/j.jecp.2010.09.008

Rachlin, H. (2002). Altruism and selfishness. *Behavioral and Brain Sciences*, *25*(2), 239–250. https://doi.org/10.1017/S0140525X02000055

Ramani, G. B., Jaeggi, S. M., Daubert, E. N., & Buschkuehl, M. (2017). Domain-specific and domain-general training to improve kindergarten children's mathematics. *Journal of Numerical Cognition*, *3*(2), 468–495. https://doi.org/10.5964/jnc.v3i2.31

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430. https://doi.org/10.1038/nature11467

Randerath, J., Valyear, K. F., Philip, B. A., & Frey, S. H. (2017). Contributions of the parietal cortex to increased efficiency of planning-based action selection. *Neuropsychologia*, *105*(July 2016), 135–143. https://doi.org/10.1016/j.neuropsychologia.2017.04.024

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, *26*(6), 462–483. https://doi.org/10.1016/j.tics.2022.03.007

Rawls, J. (1971). A theory of justice. In G. Sher (Ed.), *Ethics: Essential Readings in Moral Theory*. Cambridge (Mass.).

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment*, *31*(12), 1395–1411. https://doi.org/10.1037/pas0000754

Richmond, L., Redick, T. S., & Braver, T. S. (2015). Remembering to prepare: The benefits (and costs) of high working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1764.

Rodrigue, K. M., & Raz, N. (2004). Shrinkage of the Entorhinal Cortex over Five Years Predicts Memory Performance in Healthy Adults. *Journal of Neuroscience*, *24*(4), 956–963. https://doi.org/10.1523/JNEUROSCI.4166-03.2004

Rosseel, Y. (2012). An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Ruel, A., Devine, S., & Eppinger, B. (2021). Resource-rational approach to meta-control problems across the lifespan. In *Wiley Interdisciplinary Reviews: Cognitive Science* (Vol. 12, Issue 5). John Wiley and Sons Inc. https://doi.org/10.1002/wcs.1556

Rustichini, A. (2015). The role of intelligence in economic decision making. *Current Opinion in Behavioral Sciences*, *5*, 32–36. https://doi.org/10.1016/j.cobeha.2015.07.002

Sala, G., & Gobet, F. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review*, *18*, 46–57. https://doi.org/10.1016/j.edurev.2016.02.002

Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, *53*(4), 671–685. https://doi.org/10.1037/dev0000265

Sala, G., & Gobet, F. (2019). Cognitive Training Does Not Enhance General Cognition. *Trends in Cognitive Sciences*, *23*(1), 9–20. https://doi.org/10.1016/j.tics.2018.10.004

Sambrook, T. D., Hardwick, B., Wills, A. J., & Goslin, J. (2018). Model-free and model-based reward prediction errors in EEG. *NeuroImage*, *178*(March), 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*, 53–62. https://doi.org/10.1196/annals.1399.007

Santamaría-García, H., González-Gadea, M. L., di Tella, R., Ibáñez, A., & Sigman, M. (2018). The interplay between sharing behavior and beliefs about others in children during dictator games. *Journal of Experimental Child Psychology*, *166*, 451–464. https://doi.org/10.1016/j.jecp.2017.08.016

Sasse, L. K., Peters, J., & Brassen, S. (2017). Cognitive control modulates effects of episodic simulation on delay discounting in aging. *Frontiers in Aging Neuroscience*, *9*(MAR), 1–11. https://doi.org/10.3389/fnagi.2017.00058

Satterthwaite, T. D., Wolf, D. H., Erus, G., Ruparel, K., Elliott, M. A., Gennatas, E. D., Hopson, R., Jackson, C., Prabhakaran, K., Bilker, W. B., Calkins, M. E., Loughead, J., Smith, A., Roalf, D. R., Hakonarson, H., Verma, R., Davatzikos, C., Gur, R. C., & Gur, R. E. (2013). Functional maturation of the executive system during adolescence. *Journal of Neuroscience*, *33*(41), 16249–16261. https://doi.org/10.1523/JNEUROSCI.2345-13.2013

Schmidt, M. F. H., & Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0023223

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, *2*(JUL), 1–10. https://doi.org/10.3389/fnagi.2010.00027

Scionti, N., Cavallero, M., Zogmaister, C., & Marzocchi, G. M. (2020). Is Cognitive Training Effective for Improving Executive Functions in Preschoolers? A Systematic Review and Meta-Analysis. *Frontiers in Psychology*, *10*(January). https://doi.org/10.3389/fpsyg.2019.02812

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

Shawn Green, C., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., Bingel, U., Chein, J. M., Colzato, L. S., Edwards, J. D., Facoetti, A., Gazzaley, A., Gathercole, S. E., Ghisletta, P., Gori, S., Granic, I., Hillman, C. H., Hommel, B., Jaeggi, S. M., … Witt, C. M. (2019). Improving Methodological Standards in Behavioral Interventions for Cognitive Enhancement. *Journal of Cognitive Enhancement*, *3*(1), 2–29. https://doi.org/10.1007/s41465-018-0115-y

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shohamy, D., & Daw, N. D. (2015). Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences*, *5*, 85–90. https://doi.org/10.1016/j.cobeha.2015.08.010

Silva Castillo, L. H. (2017). Temporal Discounting and Health Behavior: A Review. *MOJ Public Health*, *6*(6). https://doi.org/10.15406/mojph.2017.06.00189

Simon, H. A. (1957). *Models of man: social and rational: mathematical essays on rational human behavior in a social setting*. Wiley.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do “Brain-Training” Programs Work? *Psychological Science in the Public Interest, Supplement*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Slavic, P., Fischhaff, B., & Lichtenstein, S. (1977). Behavioral Decision Theory. *Annual Review of Psychology*, *28*, 1–39.

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a Science of Effective Cognitive Training. *Current Directions in Psychological Science*, *29*(6), 531–537. https://doi.org/10.1177/0963721420951599

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

Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I Should but I Won't: Why Young Children Endorse Norms of Fair Sharing but Do Not Follow Them. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059510

Smithers, L. G., Sawyer, A. C. P., Chittleborough, C. R., Davies, N. M., Davey Smith, G., & Lynch, J. W. (2018). A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nature Human Behaviour*, *2*(11), 867–880. https://doi.org/10.1038/s41562-018-0461-x

Smittenaar, P., FitzGerald, T. H. B., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of Dorsolateral Prefrontal Cortex Decreases Model-Based in Favor of Model-free Control in Humans. *Neuron*, *80*(4), 914–919. https://doi.org/10.1016/j.neuron.2013.08.009

Southgate, V., Begus, K., Lloyd-Fox, S., di Gangi, V., & Hamilton, A. (2014). Goal representation in the infant brain. *NeuroImage*, *85*, 294–301. https://doi.org/10.1016/j.neuroimage.2013.08.043

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–726. https://doi.org/10.1017/S0140525X00003435

Stanovich, K. E., & West, R. F. (2003). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *26*(4), 527. https://doi.org/10.1017/S0140525X03210116

Steinbeis, N. (2016). Taxing behavioral control diminishes sharing and costly punishment in childhood. *Developmental Science*, *21*(1), 1–8. https://doi.org/10.1111/desc.12492

Steinbeis, N. (2018). Neurocognitive mechanisms of prosociality in childhood. *Current Opinion in Psychology*, *20*, 30–34. https://doi.org/10.1016/j.copsyc.2017.08.012

Steinbeis, N., Bernhardt, B. C., & Singer, T. (2012). Impulse Control and Underlying Functions of the Left DLPFC Mediate Age-Related and Age-Independent Individual Differences in Strategic Social Behavior. *Neuron*, *73*(5), 1040–1051. https://doi.org/10.1016/j.neuron.2011.12.027

Steinbeis, N., Haushofer, J., Fehr, E., & Singer, T. (2016). Development of Behavioral Control and Associated vmPFC-DLPFC Connectivity Explains Children's Increased Resistance to Temptation in Intertemporal Choice. *Cerebral Cortex*, *26*(1), 32–42. https://doi.org/10.1093/cercor/bhu167

Steinbeis, N., & Over, H. (2017). Enhancing behavioral control increases sharing in children. *Journal of Experimental Child Psychology*, *159*, 310–318. https://doi.org/10.1016/j.jecp.2017.02.001

Steiner, M. D., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, *5*(53), 2521. https://doi.org/10.21105/joss.02521

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Stevens, M. C., Skudlarski, P., Pearlson, G. D., & Calhoun, V. D. (2009). Age-related cognitive gains are mediated by the effects of white matter development on brain network integration. *NeuroImage*, *48*(4), 738–746. https://doi.org/10.1016/j.neuroimage.2009.06.065

Story, G. W., Vlaev, I., Seymour, B., Darzi, A., & Dolan, R. J. (2014). Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Frontiers in Behavioral Neuroscience*, *8*(MAR), 1–20. https://doi.org/10.3389/fnbeh.2014.00076

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning : an introduction* (second edition). the MIT Press.

Svetlova, M., Nichols, S. R., & Brownell, C. A. (2010). Toddlers' Prosocial Behavior: From Instrumental to Empathic to Altruistic Helping. *Child Development*, *81*(6), 1814–1827. https://doi.org/10.1111/j.1467-8624.2010.01512.x

Tate, L. M. (2015). Temporal discounting rates and their relation to exercise behavior in older adults. *Physiology and Behavior*, *152*, 295–299. https://doi.org/10.1016/j.physbeh.2015.10.003

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, *112*(1). https://doi.org/10.1086/380085

Thompson, C., Barresi, J., & Moore, C. (1997). The development of future-oriented prudence and altruism in preschoolers. *Cognitive Development*, *12*(2), 199–212.

Thorndike, E. L. (1911). *Animal intelligence*. The MacMillan company.

Tobias, S., & Carlson, J. E. (1969). Brief report: Bartlett's test of sphericity and chance findings in factor analysis. *Multivariate Behavioral Research*, *4*(3), 375–377. https://doi.org/10.1207/s15327906mbr0403_8

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, *90*(4), 293.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*(1), 67–81. https://doi.org/10.1016/j.intell.2004.08.003

Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Verbruggen, F., Adams, R., & Chambers, C. D. (2012). Proactive Motor Control Reduces Monetary Risk Taking in Gambling. *Psychological Science*, *23*(7), 805–815. https://doi.org/10.1177/0956797611434538

Verbruggen, F., Aron, A. R., Band, G. P. H., Beste, C., Bissett, P. G., Brockett, A. T., Brown, J. W., Chamberlain, S. R., Chambers, C. D., Colonius, H., Colzato, L. S., Corneil, B. D., Coxon, J. P., Dupuis, A., Eagle, D. M., Garavan, H., Greenhouse, I., Heathcote, A., Huster, R. J., … Boehler, C. N. (2019). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *ELife*, *8*, 1–26. https://doi.org/10.7554/eLife.46323

Veselic, S., Smid, C. R., & Steinbeis, N. (2021). Developmental changes in reward processing are reward specific. *PsyArXiv*, 1–36.

Wan Lee, S., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, *81*(3), 687–699. https://doi.org/10.1016/j.neuron.2013.11.028

Wang, W., & Liu, W. (2022). *Explaining Reciprocal Sharing in Repeated Dictator Game : The Role of Theory of Mind*. 100–115. https://doi.org/10.4236/jss.2022.107009

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, *311*(5765), 1301–1303. https://doi.org/10.1126/science.1121448

Wass, S. v., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory - Is younger, better? In *Developmental Review*. https://doi.org/10.1016/j.dr.2012.07.001

Weafer, J., Baggott, M. J., & de Wit, H. (2013). Test-retest reliability of behavioral measures of impulsive choice, impulsive action, and inattention.

Experimental and Clinical Psychopharmacology, *21*(6), 475–481. https://doi.org/10.1037/a0033659

Weber, B. J., & Huettel, S. A. (2008). The neural substrates of probabilistic and intertemporal decision making. *Brain Research*, *1234*, 104–115. https://doi.org/10.1016/j.brainres.2008.07.105

Wechsler, D. (2011). *WASI-II: Wechsler abbreviated scale of intelligence*.

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*, *January*, 1–10. https://doi.org/10.1111/desc.13088

Wesley, M. J., & Bickel, W. K. (2014). Remember the future II: Meta-analyses and functional overlap of working memory and delay discounting. *Biological Psychiatry*, *75*(6), 435–448. https://doi.org/10.1016/j.biopsych.2013.08.008

White, N., Forsyth, B., Lee, A., & Machado, L. (2018). Repeated Computerized Cognitive Testing: Performance Shifts and Test-Retest Reliability in Healthy Young Adults. *Psychological Assessment*, *30*(4), 539–549. https://doi.org/10.1037/pas0000503.supp

Wiebe, S. A., & Karbach, J. (2017). Executive function: Development across the life span. In *Executive Function: Development Across the Life Span*. https://doi.org/10.4324/9781315160719

Wilkinson, H. R., Smid, C., Morris, S., Farran, E. K., Dumontheil, I., Mayer, S., Tolmie, A., Bell, D., Porayska-Pomsta, K., Holmes, W., Mareschal, D., & Thomas, M. S. C. (2019). Domain-Specific Inhibitory Control Training to Improve Children's Learning of Counterintuitive Concepts in Mathematics and Science. *Journal of Cognitive Enhancement*. https://doi.org/10.1007/s41465-019-00161-4

Williams, B. R., Strauss, E. H., Hultsch, D. F., & Hunter, M. A. (2007). Reaction time inconsistency in a spatial stroop task: Age-related differences through childhood and adulthood. *Aging, Neuropsychology, and Cognition*, *14*(4), 417–439. https://doi.org/10.1080/13825580600584590

Worsley, K. J., Taylor, J. E., Carbonell, F., Chung, M. K., Duerden, E., & Bernhardt, B. (2009). *SurfStat. a Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory* (p. 47).

Yong, A. G., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.p079

Yoon, U., Lee, J. M., Im, K., Shin, Y. W., Cho, B. H., Kim, I. Y., Kwon, J. S., & Kim, S. I. (2007). Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *NeuroImage*, *34*(4), 1405–1415. https://doi.org/10.1016/j.neuroimage.2006.11.021

Zaki, J., & Mitchell, J. P. (2013). Intuitive Prosociality. *Current Directions in Psychological Science*, *22*(6), 466–470. https://doi.org/10.1177/0963721413492764

Zhang, R., & Wang, Z. (2019). The mediating effect of empathy in the quadratic relationship between children's resting RSA and sharing behavior. *International Journal of Psychophysiology*, *140*(199), 8–14. https://doi.org/10.1016/j.ijpsycho.2019.03.012

Zhao, X., Wang, Y., & Maes, J. H. R. (2022). The effect of working memory capacity and training on intertemporal decision making in children from low-socioeconomic-status families. *Journal of Experimental Child Psychology*, *216*, 105347. https://doi.org/10.1016/j.jecp.2021.105347

7. Author contribution statement

The computational model included in experimental Chapters 2 and 3 was based on the original dual-systems reinforcement learning model reported in Daw et al. 2011 and the current computational model included in the thesis is based on work conducted by Wouter Kool and Sam Gershman (Daw et al., 2011; Kool et al., 2016). The MRI data collected and reported in this thesis in experimental Chapter 3 was collected with the help of Dr. Abigail Thompson and Keertana Ganesan. Keertana Ganesan conducted the confirmatory factor analysis for the executive functions reported in Chapter 4.

8. Appendices

8.1 Published work

Attached at the end of this section are two papers: the experimental paper based on work conducted in Chapter 2 and the review paper cited in Chapters 4 and 5.

Below is a complete list of published works during the Ph.D.:

Cañigueral, R., Ganesan, K., Smid, C. R., Thompson, A., Dosenbach, N. U., & Steinbeis, N. (2022). Adaptiveness of fluctuations in intra-individual variability of performance is process-dependent in middle childhood. *PsyArXiv*. https://doi.org/10.31234/osf.io/y7c5d **[Preprint]**

Ganesan, K., Smid, C. R., Thompson, A., Buchberger, E. S., Spowage, J., Iqbal, S., ... & Steinbeis, N. (2021). Not context monitoring, but inhibition plays a privileged role in childhood cognitive control. *PsyArXiv*. https://doi.org/10.31234/osf.io/kuebx **[Preprint]**

Kentrop, J., Smid, C. R., Achterberg, E. J. M., Van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Joëls, M., & Van der Veen, R. (2018). Effects of maternal deprivation and complex housing on rat social behavior in adolescence and adulthood. *Frontiers in behavioral neuroscience*, *12*, 193. https://doi.org/10.3389/fnbeh.2018.00193

Steinbeis, N., & Smid, C. R. (2022). Interventions for Improving Executive Functions during Development: Working Memory, Cognitive Flexibility, and Inhibition. In O. Houdé & G. Borst (Eds.), The Cambridge Handbook of Cognitive Development (Cambridge Handbooks in

Psychology, pp. 623-643). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108399838

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a science of effective cognitive training. *Current Directions in Psychological Science*, *29*(6), 531-537. https://doi.org/10.1177%2F0963721420951599

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/10.1111/desc.13295

Veselic, S., Smid, C. R., & Steinbeis, N. (2021). Developmental changes in reward processing are reward specific. *PsyArXiv*. https://doi.org/10.31234/osf.io/fzk9t **[Preprint]**

Wilkinson, H. R., Smid, C., Morris, S., Farran, E. K., Dumontheil, I., Mayer, S., ... & Thomas, M. S. (2020). Domain-specific inhibitory control training to improve children's learning of counterintuitive concepts in mathematics and science. *Journal of Cognitive Enhancement*, *4*(3), 296-314. https://doi.org/10.1007/s41465-019-00161-4

Received: 10 October 2021 | Revised: 12 May 2022 | Accepted: 22 May 2022

DOI: 10.1111/desc.13295

Developmental Science WILEY

RESEARCH ARTICLE

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

Claire R. Smid[1] | Wouter Kool[2] | Tobias U. Hauser[3,4] | Nikolaus Steinbeis[1]

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

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

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

[4]Wellcome 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 computationally 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.

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 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.

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

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,

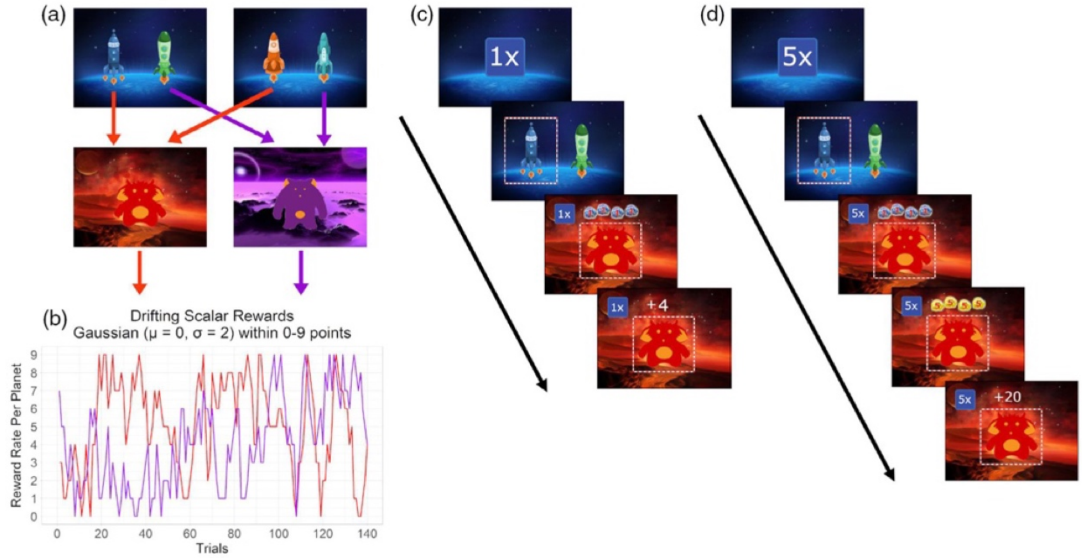


FIGURE 1 *Task Design.* (a) Schematic of the transition structure with arrows displaying deterministic transitions; if a participant chose the dark blue or the orange spaceship, they would always transition to the red planet. (b) At the planets, participants received rewards in the form of space treasure 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 (Lüdecke, 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

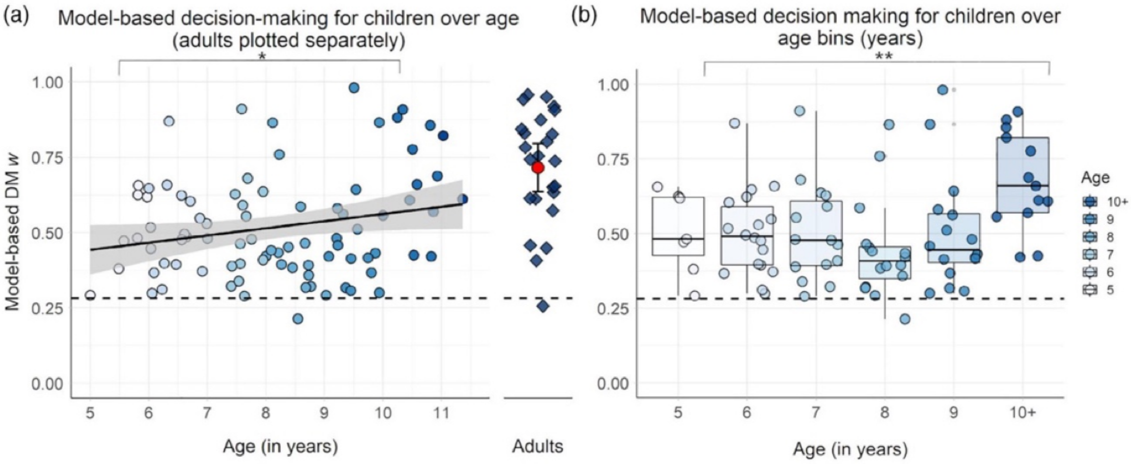


FIGURE 2 *Model-based decision-making over age for children with the simulated model-free baseline.* (a) The degree of model-based decision-making significantly increased with age for the children. The dashed line represents the grand mean of the model-free simulations, which acts as the simulated model-free 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

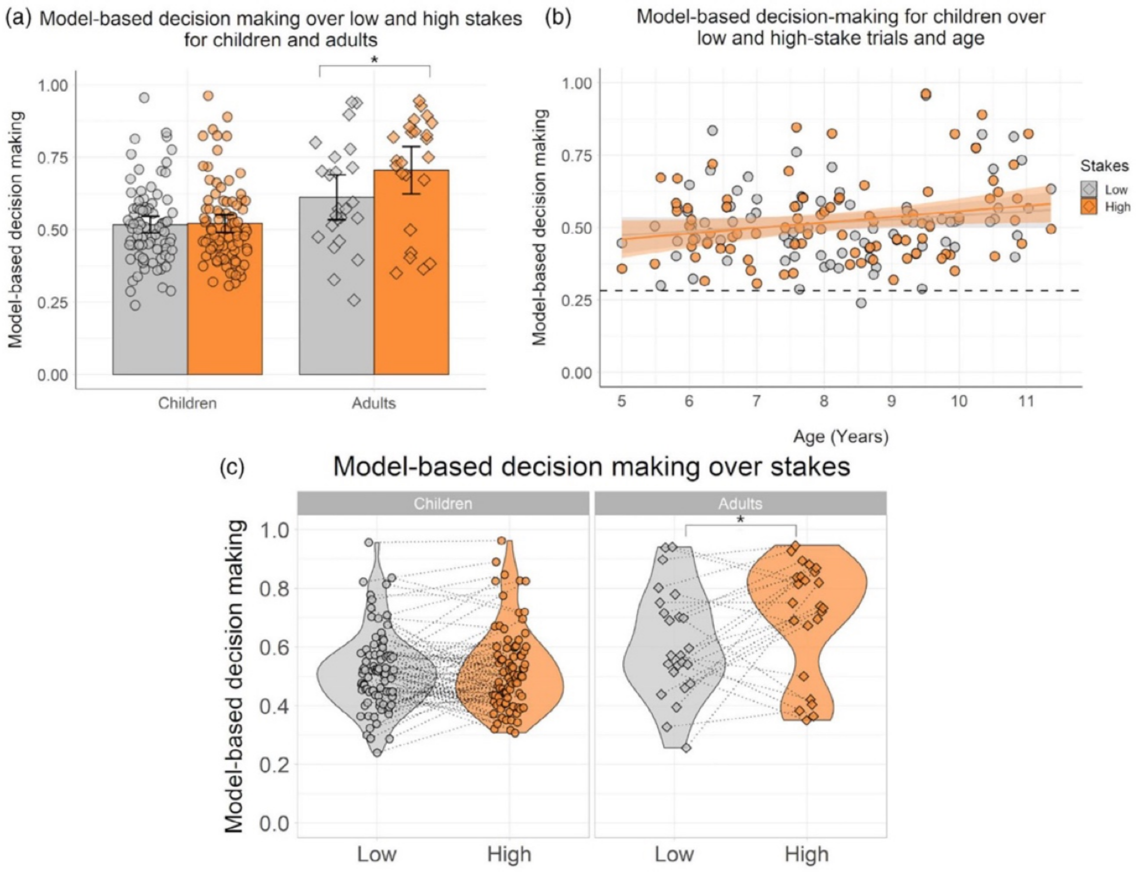


FIGURE 3 *Model-based decision-making over stakes for adults and children.* (a) Adults displayed a significantly higher degree of model-based decision-making for the high-stake trials, while children did not show a difference in the degree of model-based decision-making used over stakes. (b) this did 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

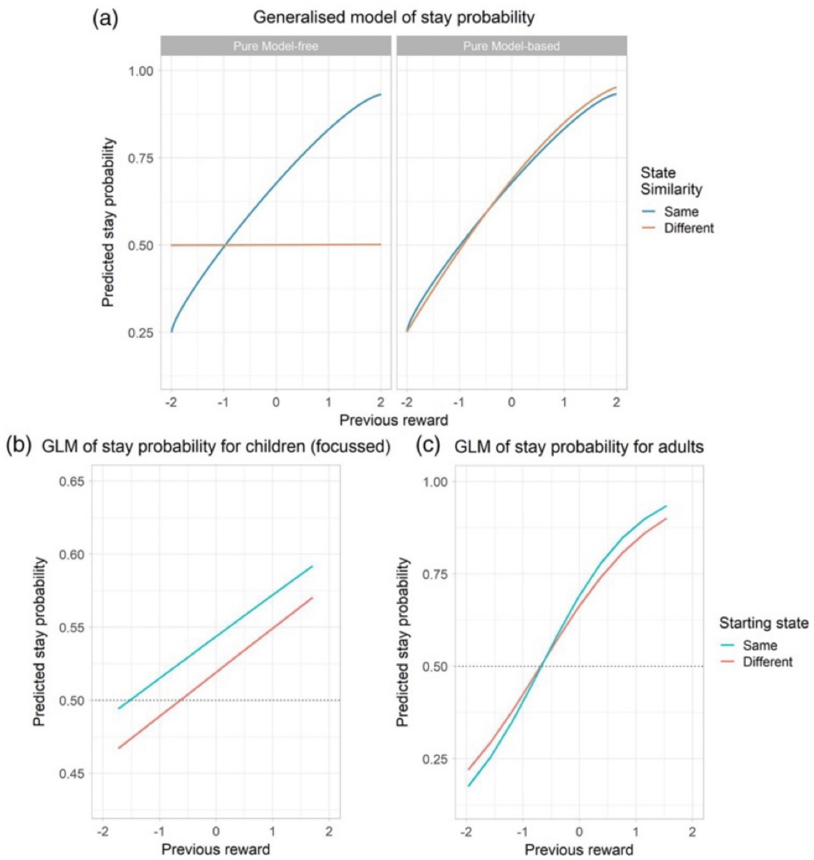


FIGURE 4 *Model-free and model-based contributions to stay probability*. Stay probability meant repeating a visit to the same planet (red or purple, see Figure 1a). (a) Examples of influences of pure model-free and model-based decision making on stay probability following previous reward. For a pure 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

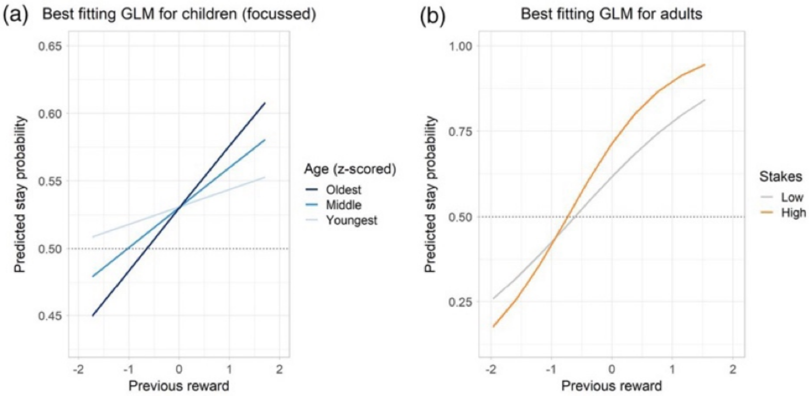


FIGURE 5 *Best fitting generalized linear mixed models of stay probability for the children and adults.* Stay probability meant repeating a visit to the same planet (red or purple, see Figure 1a). (a) Predicted results from the best-fitting model for children. Previous reward—the model-based component—was 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 (β = 0.12, se = 0.02, z = 5.60, p < 0.001), while age was not a significant main effect (β = -0.00, se = 0.04, z = -0.04, p = 0.967), but the interaction between previous reward and age was significant (β = 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) (β = 1.14, se = 0.05, z = 22.78, p < 0.001) and stake (β = 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 (β = 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 (β = 0.13, se = 0.05, z = 2.56, p = 0.010), and the three-way interaction between previous points, starting state and stake (β = 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, (β = 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 (β = 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 metacontrol, 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., & Lüdecke, 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

Lüdecke, 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). *AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c)* (R package version 2.3-1). https://cran.r-project.org/package=AICcmodavg

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

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
2020, Vol. 29(6) 531–537

DOI: 10.1177/0963721420951599
www.psychologicalscience.org/CDPS
SAGE

Toward a Science of Effective Cognitive Training

Claire R. Smid[1], Julia Karbach[2,3], and Nikolaus Steinbeis[1]
[1]Department of Clinical, Educational and Health Psychology, University College London;
[2]Department of Psychology, University of Koblenz-Landau; and [3]Center for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt, Germany

Abstract
A long-standing question in the behavioral sciences is whether cognitive functions can be improved through dedicated training. It is uncontested that training programs can lead to *near transfer*, meaning increased performance on untrained tasks involving similar cognitive functions. However, whether training also leads to *far transfer*, meaning increased performance on loosely related untrained tasks or even activities of daily living, is still hotly debated. Here, we review the extant literature and, in particular, the most recent meta-analytic evidence and argue that the ongoing crisis in the field of cognitive-training research may benefit from taking a more mechanistic approach to studying the effectiveness of training. We propose that (a) adopting a more rigorous theoretical framework that builds on a process-based account of training and transfer, (b) considering the role of individual differences in the responsiveness to training, and (c) drawing on Bayesian models of development may help to solve controversial issues in the field and lead the way to designing and implementing more effective training protocols.

Keywords
cognitive training, transfer effects, cognitive development, mechanisms, Bayesian learning, computational modeling

Practice supposedly makes perfect. But does it also lead to tangible improvements in skills not directly trained? Addressing how experiences generalize beyond the context in which they take place can answer fundamental questions of cognitive architecture and learning. Since the early 2000s, there has been a particular interest in whether executive functions (EFs) can be improved (Smithers et al., 2018), in particular, working memory (WM), inhibitory control (IC), and cognitive flexibility (CF; for a review, see Strobach & Karbach, 2021). This interest was nurtured by findings that EFs in childhood are linked to academic achievement, mental health, social functioning, and well-being both during childhood and especially later in life (Moffitt et al., 2011). As a result, attempts to impinge on these critical life skills have surged, but findings remain equivocal (Diamond & Ling, 2020; Redick, 2019; Titz & Karbach, 2014).

The "brain-training industry" has capitalized on man's tireless endeavor to self-improve, as indicated by a forecasted net worth of more than $8 billion by 2021 (Ahuja, 2019). This mandates a critical examination of the quality of existing evidence for the benefits of cognitive training of EFs against stringent criteria. Despite several recent best-practices recommendations (Green et al., 2019; Simons et al., 2016) for evaluating the effectiveness of cognitive training, a comprehensive understanding of how, for whom, and why certain training can be effective is still missing. Thus, we need a mechanistic framework on cognitive training, integrating methodology and theory to drive the field forward.

We briefly review theoretical positions and empirical evidence in favor of and against the effectiveness of EF training. What emerges is a striking discord in the field, with strong claims and supporting evidence on both sides, giving rise to the questions of how and whether these discrepancies can be reconciled. We propose three key paradigm shifts to facilitate a rapprochement and suggest novel and necessary ways to assess whether and how cognitive training can be effective: (a) establishing

Corresponding Author:
Nikolaus Steinbeis, University College London, Division of Psychology and Language Sciences, London, WC1H 0AP, United Kingdom
E-mail: n.steinbeis@ucl.ac.uk

a mechanistic link between a training mechanism and a transfer domain, (b) considering the importance of individual differences in the effectiveness of training, and (c) offering a theoretical perspective on when and how particular training interventions might be most effective. We hope that this will facilitate incorporating changes into both training design and analysis and clarify how training might impact cognitive functions.

What Is the Consensus on Cognitive Training?

Although it is uncontested that training can impact closely related domains (*near transfer*), it is still intensely debated whether they lead to improvements in loosely related domains (*far transfer*; Diamond & Ling, 2020). Theories on the possibility of far transfer also range in their optimism. In their *common-elements theory*, Thorndike and Woodworth (1901) argued that transfer happens within one domain via knowledge that shares common elements but that far transfer is rare. Since then, Anderson (1982) assumed that *production rules* coordinate exchange between specialized cognitive systems but are often specific to a particular task. In contrast, the *primitive-information-processing-elements theory* (Taatgen, 2013) claims that training on particular tasks evolves a set of operators toward that task, which should be useful in new contexts and lead to transfer. The cognitive-routine framework (Gathercole, Dunning, Holmes, & Norris, 2019) posited that participants develop new cognitive routines during training. These routines are automated cognitive procedures that can be applied to novel tasks sharing the same requirements. Transfer to other tasks will occur if there are common task features (e.g., transfer of WM training to IC after training on complex but not simple span tasks). One drawback is that available models of transfer make only very general and limited predictions about the conditions under which far transfer should occur.

Recent meta-analyses on the effectiveness of training reflect these diverse theories. Sala and Gobet (2017, 2019) provided a critical assessment of transfer effects after WM training. They concluded that cognitive training does not enhance general cognition because effect sizes for far transfer are low, and effect size is inversely related to the quality of study design. In contrast, a meta-analysis including numerous cognitive-training interventions (EF training, classroom-based and game-based activities) demonstrated far transfer across a wide range of domains (literacy, numeracy, language skills, IQ, and psychosocial outcomes; Smithers et al., 2018). These findings echo previous observations that highly contextualized training is most likely to yield transfer effects (Diamond & Lee, 2011). Further, Smithers et al. (2018) found that better quality studies yielded larger effect sizes, again in contradiction to Sala and Gobet's conclusions. Other studies imply that nonspecific training interventions seem to generate more generalized outcomes (Heckman, Pinto, & Savelyev, 2013; Lillard & Else-Quest, 2006), suggesting that more holistic programs, including multidimensional content, better support overall child development and yield broad-based benefits. Below, we offer a theoretical perspective for why this could be the case.

Mechanisms: Establishing a Framework

A core assumption of training studies is that training mechanisms are fundamentally related to outcome measures of interest (for a review, see Noack, Lovden, Schmiedek, & Lindenberger, 2009). For instance, WM capacity (i.e., the maximal amount of information that can be stored and manipulated in WM) correlates highly with general intelligence (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). The logical and empirical consequence has been to target WM capacity to increase intelligence. However, as has been argued elsewhere, two correlated variables, such as WM span and fluid reasoning, do not necessarily covary when one is being artificially inflated through training, because training can tap unshared variance between the two constructs (Moreau & Conway, 2014). Moreover, although relationships between WM and intelligence might exist at a latent factor level, this is not necessarily the case at the level of single tasks that are typically used in training studies. Also, EFs are higher-order constructs including different processes. For instance, WM consists of storage, rehearsal, and matching as well as manipulation of information and processing skill. Correlating two tasks does not offer sufficient granularity or direction to identify the true underlying process-based nature of the relationship. Finally, considering task manipulations (i.e., increasing WM span) as tantamount to training for outcome variables is a nontrivial endeavor. For example, it has been shown that it is not WM span per se that is related to intelligence (Unsworth & Engle, 2005) but rather a shared executive attention-control mechanism required for the active maintenance of information in the face of concurrent processing and interference. Increasing WM span may therefore not do much to improve intelligence (Sala & Gobet, 2017).

To remedy these shortcomings, we propose the following: First, we need to understand the true relationship between training mechanisms and outcome variables. This is a challenging endeavor for many reasons, among them the *task-impurity problem* in the measurement of EFs (Kane & Engle, 2003; Miyake & Friedman, 2012). Although much progress is being made

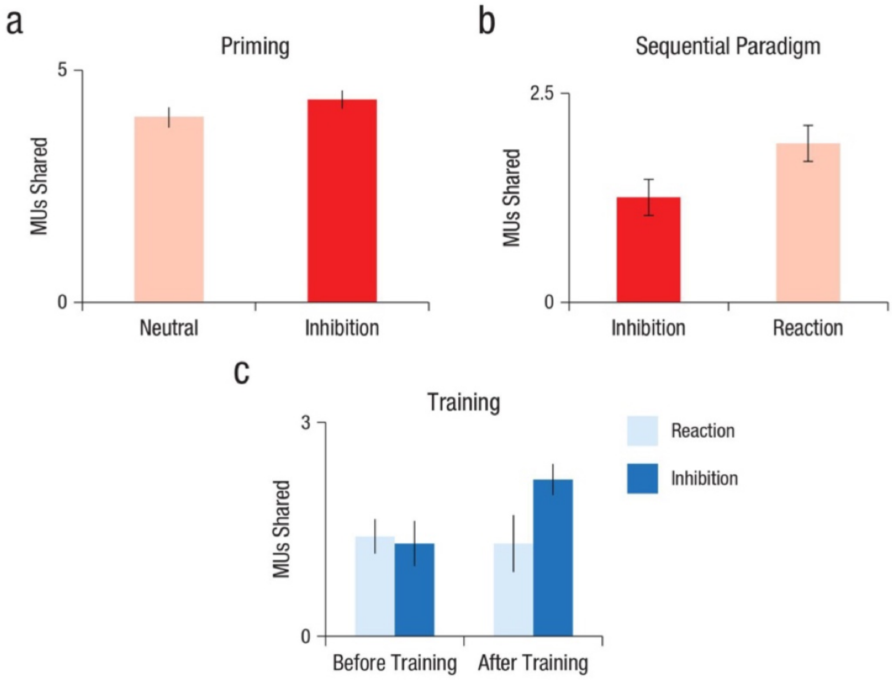


Fig. 1. Effects of different inhibitory-control manipulations (i.e., priming, sequential paradigm, training) on prosocial behavior (i.e., sharing monetary units, or MUs, in a dictator game). Compared with neutral priming, priming inhibition increases sharing (a). Compared with a reaction task, an inhibition task subsequently decreases sharing (b). Compared with training reaction speed, training inhibition increases sharing (c). Error bars show standard errors.

using latent-variable approaches (Könen & Auerswald, 2021), additional approaches may contribute to our understanding of the mechanisms underlying training and transfer effects. Generative computational models allow the parsing of task performance into multiple distinct processes that necessitate different computations as well as into directionality between processes (Sutton & Barto, 2018). Recently, computational models have elucidated processes underlying WM performance (time-based resource-sharing models; Oberauer & Lewandowsky, 2011) and IC (Bayesian ex-Gaussian estimation of reaction times; Matzke, Dolan, Logan, Brown, & Wagenmakers, 2013). For instance, canonical analyses of standard inhibition tasks such as the stop-signal response time task provide a single measure of mean performance. Recent computational frameworks using Bayesian ex-Gaussian estimation of reaction times decompose the signal into μ and σ parameters, which are the mean and standard deviation of the Gaussian component, whereas τ reflects the tail of the distribution (Matzke et al., 2013). It has been argued that whereas mean performance indicates inhibitory capacity, the tail indicates lapses of attention (Schel, Thompson, & Steinbeis, 2020). This can be used to inform the design of cognitive training targeting constituent processes of core EFs. Computational modeling thus offers promise in identifying which training mechanisms need to be targeted to affect specific outcome variables.

Second, to ascertain that appropriate training mechanisms are identified and targeted, we propose to draw on experimental manipulations and not correlations. Experimental manipulations such as dual-task paradigms or priming studies offer a means to establish mechanistic relationships between variables without the cost of full-fledged training interventions. One way to manipulate EFs is to apply dual-task or serial-task paradigms. For example, it has been argued that IC is required to overcome the temptation of keeping resources for oneself and share with anonymous others (Steinbeis, 2018a). After showing that manipulating IC impacts prosocial behavior (not through training but through two lower-cost serial-task paradigms; Figs. 1a and 1b, respectively; Steinbeis, 2018b; Steinbeis & Over, 2017), we could proceed to train IC and test whether this leads to direct changes in prosocial behavior (Steinbeis, 2020; see Fig. 1c). Third, training needs to be delivered across a range of tasks and not just single manifestations in order to show change on the (latent) ability level (Noack, Lovden, & Schmiedek, 2014). Fourth, an appropriate operationalization of training mechanisms in question is required. For instance, IC training often simply reduces the response time window (Enge et al., 2014; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009), which might train the speed at which a capacity is deployed but not capacity itself. Much more careful thought needs to be given to how capacities, rather than just task performance, can be improved.

We therefore propose, where possible, to (a) employ computational models allowing constructs to be understood in terms of constituent processes, (b) systematically manipulate these processes in dual-task frameworks to assess whether they have an impact on relevant outcome measures, and (c) carefully consider, once a training mechanism has been identified, how it can be manipulated to train the desired outcome variable.

Individual Differences: Personalizing Training

State-of-the-art studies consistently show that individuals respond differently to the same training intervention. These interindividual differences in training gains often show distinct patterns after different types of training, with compensation effects (larger gains in low performers) particularly emerging after process-based training targeting one or more basic cognitive resources (e.g., EFs) and magnification effects (larger gains in high performers) typically appearing after strategy-based training, such as mnemonics (Karbach & Verhaeghen, 2014). Training-induced gains also vary as a function of individual differences in such factors as age, baseline ability, motivation, personality, and genetic predisposition (Strobach & Karbach, 2021), indicating that especially low-performing and at-risk individuals can benefit massively from EF training (Karbach, Könen, & Spengler, 2017). And yet these individual differences are often overlooked, and current approaches broadly take univariate statistical approaches, which are unable to identify individual cognitive profiles of performance on the basis of rich multivariate data. Contemporary multivariate analysis methods offer a radical rethink of training and associated transfer (Astle, Bathelt, The CALM Team, & Holmes, 2019; Bathelt, Holmes, & Astle, on behalf of the CALM Team, 2018) by focusing on training-related changes in task relationships.

Moreover, we need to consider intraindividual dynamics in training-related performance changes. Intraindividual performance trajectories across training reveal which participants show training effects and when they reach their individual maximum. The fluctuations in these trajectories can be indicative of adaptive processes (e.g., varying strategies) or maladaptive processes (e.g., vulnerability to disturbing influences). Intraindividual couplings of performance fluctuations with other variables (e.g., motivation, affect) can tell us which internal and external factors contribute to individual performances and to what extent participants differ in the strength of these relations (Könen & Karbach, 2015). This seems particularly relevant for studies with heterogeneous samples because variation in intraindividual effects across training may eventually result in interindividual differences in training outcomes. Considering both inter- and intraindividual differences and dynamics is likely to contribute massively to our understanding of training outcomes and can help generate theories regarding the underlying mechanisms.

Finally, these considerations may also help to explain the heterogeneous findings that extend to the level of meta-analyses: Looking at mean group changes in primary studies and averaging across their effect sizes in meta-analyses does not do justice to interindividual and intraindividual differences. We therefore propose to investigate who benefits the most in order to design tailored training studies targeting EFs and the numerous outcomes building on them. As in fields of medicine, which have embraced the necessity of personalizing treatment, researchers in the field of cognitive training need to consider differences in variables such as baseline ability, motivation and affect, genetic predisposition, environmental experience, and lifestyle as well as developmental stage and individually and developmentally relevant goals.

Theory: Bayesian Account of Development

Current theoretical accounts on far transfer lack prediction on what a training intervention must entail to be effective. The interactive-specialization hypothesis on brain development states that cortical circuits specialize over development (Johnson, 2001, 2011) and that training should be particularly effective during childhood (Wass, Scerif, & Johnson, 2012). We argue that a more fine-grained definition of how the training input is perceived is critical for understanding whether transfer occurs. Bayesian learning accounts have recently been used to study developmental plasticity (Fawcett & Frankenhuis, 2015; Stamps & Frankenhuis, 2016). Bayes's theorem provides the most logically consistent way to model an individual's current assessment of conditions in the external environment (the state of the world) using a probability distribution. Such models assume that individuals have naive priors, which are updated as individuals are exposed to a series of potentially informative cues over the course of their lives, yielding a series of posterior distributions.

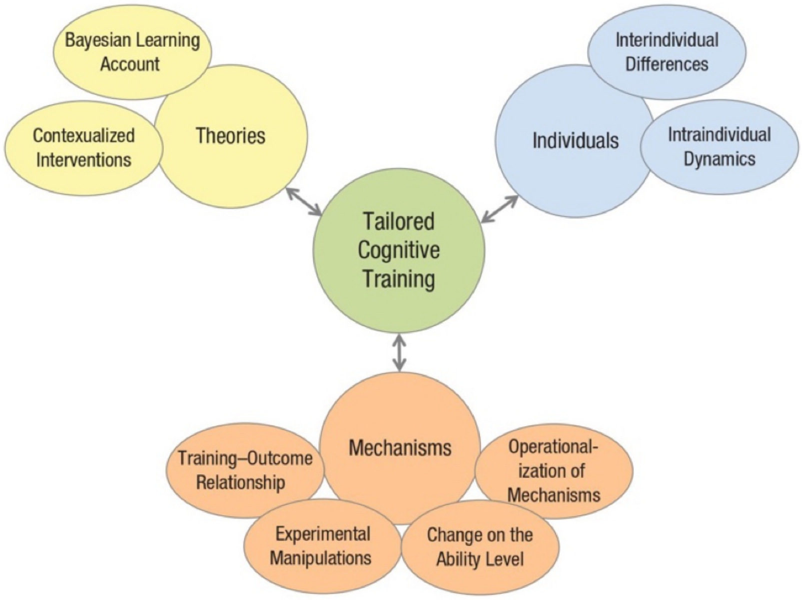


Fig. 2. Illustration of the proposed framework for tailored cognitive training derived from a process-based theoretical account (mechanisms). Drawing on Bayesian models of development, the framework considers the role of individual differences in the responsiveness to interventions.

Development unfolds as a function of children's assessment of the state of their world, as reflected by their posterior distributions.

Essential to a Bayesian learning account is the processing of cues, which refers to experiences that are potentially informative about environmental conditions. Cues are primarily assessed in terms of their reliability and informativeness (however, note that cues can be uninformative, unreliable, and misleading). Cue reliability is determined by the likelihood that a specific cue will occur given each possible state of the world. Cue informativeness refers to the extent to which cues are informative by reducing uncertainty about the world (Fawcett & Frankenhuis, 2015). We suggest that cognitive training can be seen as a set of cues, which are assessed in terms of these criteria and thus how representative of ecologically meaningful events and predictive of current or future contexts they are. In sum, in the context of cognitive training, a cue is a stimulus that is informative about the world outside the context of the training. Such information could be engendered by stimuli of specific significance (e.g., highly caloric food for dieters) or by relevant contexts (e.g., learning inhibition in the context of a social interaction).

A prediction of this framework is that training with poor reliability and informativeness in terms of an individual's actual current or future experience of the world is likely to have limited to no impact. Context is therefore of particular relevance. Evidence in support of this comes precisely from the meta-analytic studies presented above, whereby isolated training focusing only on specific aspects of cognition without any embedding led to limited transfer, whereas training that was contextualized in terms of how and where it was delivered was shown to be more effective (Diamond & Lee, 2011). Interestingly, this framework can also account for effective transfer to real-life drinking and eating behavior following from highly specific training cues (i.e., inhibition of response tendencies to alcoholic drinks to reduce alcohol consumption; Jones et al., 2016). Bayesian accounts of learning offer both test bed and guidance for the design of effective training studies.

Conclusion

In this article, we have argued that the ongoing crisis in the field of cognitive-training research may benefit from taking a more fine-grained approach to training studies (Fig. 2). We argue that (a) adopting a more rigorous theoretical framework that builds on a process-based account of the underlying mechanisms of training and transfer, (b) considering the role of individual differences in the responsiveness to training, and (c) drawing on Bayesian models of development may help solve controversial issues in the field and lead the way to designing and implementing highly effective tailored training.

Recommended Reading

Bathelt, J., Holmes, J., & Astle, D. E., on behalf of the Centre for Attention Learning and Memory (CALM) Team. (2018). (See References). Presents a multivariate approach to understanding performance profiles.

Novick, J. M., Bunting, M. F., Dougherty, M. R., & Engle, R. W. (Eds.). (2020). *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development*. New York, NY: Oxford University Press. Showcases evidence on effectiveness of cognitive trainings.

Smithers, L. G., Sawyer, A. C. P., Chittleborough, C. R., Davies, N., Smith, G. D., & Lynch, J. (2018). (See References). Presents meta-analytic evidence on far-transfer effects following cognitive interventions.

Stamps, J. A., & Frankenhuis, W. E. (2016). (See References). Provides a Bayesian model of development.

Transparency

Action Editor: Randall W. Engle
Editor: Randall W. Engle
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding
N. Steinbeis is currently funded by the Jacobs Foundation as well as the European Research Council (DEVBRAIN-TRAIN/715282).

ORCID iD

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

References

Ahuja, A. (2019, January 23). An evidence deficit haunts the billion-dollar brain training industry. *Financial Times*. Retrieved from www.ft.com/content/a0166eea-1e41-11e9-a46f-08f9738d6b2b

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.

Astle, D. E., & Bathelt, J., The CALM Team, & Holmes, J. (2019). Remapping the cognitive and neural profiles of children who struggle at school. *Developmental Science*, *22*, Article e12747. doi:10.1111/desc.12747

Bathelt, J., Holmes, J., & Astle, D. E., on behalf of the Centre for Attention Learning and Memory (CALM) Team. (2018). Data-driven subtyping of executive function-related behavioral problems in children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*, 252–262. doi:10.1016/j.jaac.2018.01.014

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*, 959–964.

Diamond, A., & Ling, D. S. (2020). Review of the evidence on, and fundamental questions about, efforts to improve executive functions, including working memory. In J. M. Novick, M. F. Bunting, M. R. Dougherty, & R. W. Engle (Eds.), *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development* (pp. 143–431). New York, NY: Oxford University Press.

Enge, S., Behnke, A., Fleischhauer, M., Kuttler, L., Kliegel, M., & Strobel, A. (2014). No evidence for true training and transfer effects after inhibitory control training in young healthy adults. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*, 987–1001.

Fawcett, T. W., & Frankenhuis, W. E. (2015). Adaptive explanations for sensitive windows in development. *Frontiers in Zoology*, *12*, Article S3. doi:10.1186/1742-9994-12-S1-S3

Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2019). Working memory training involves learning new skills. *Journal of Memory and Language*, *105*, 19–42.

Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K., . . . Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Neuroscience*, *3*, 2–29.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*, 2052–2086.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, USA*, *105*, 6829–6833.

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, *2*, 475–483.

Johnson, M. H. (2011). Interactive specialization: A domain-general framework for human functional brain development? *Developmental Cognitive Neuroscience*, *1*, 7–21.

Jones, A., Di Lemma, L. C., Robinson, E., Christiansen, P., Nolan, S., Tudur-Smith, C., & Field, M. (2016). Inhibitory control training for appetitive behaviour change: A meta-analytic investigation of mechanisms of action and moderators of effectiveness. *Appetite*, *97*, 16–28.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47–70.

Karbach, J., Könen, T., & Spengler, M. (2017). Who benefits the most? Individual differences in the transfer of executive control training across the lifespan. *Journal of Cognitive Enhancement*, *1*, 394–405.

Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, *25*, 2027–2037.

Könen, T., & Auerswald, M. (2021). Statistical modeling of latent change. In T. Strobach & J. Karbach (Eds.), *Cognitive training: An overview of features and applications* (2nd ed., pp. 55–68). Cham, Switzerland: Springer.

Könen, T., & Karbach, J. (2015). The benefits of looking at intraindividual dynamics in cognitive training data. *Frontiers in Psychology, 6*, Article 615. doi:10.3389/fpsyg.2015.00615

Lillard, A., & Else-Quest, N. (2006). The early years: Evaluating Montessori education. *Science, 313*, 1893–1894.

Matzke, D., Dolan, C. V., Logan, G. D., Brown, S. D., & Wagenmakers, E. J. (2013). Bayesian parametric estimation of stop-signal reaction time distributions. *Journal of Experimental Psychology: General, 142*, 1047–1073.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*, 8–14.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, USA, 108*, 2693–2698.

Moreau, D., & Conway, A. R. (2014). The case for an ecological approach to cognitive training. *Trends in Cognitive Sciences, 18*, 334–336.

Noack, H., Lovden, M., & Schmiedek, F. (2014). On the validity and generality of transfer effects in cognitive training research. *Psychological Research, 78*, 773–789.

Noack, H., Lovden, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. *Restorative Neurology and Neuroscience, 27*, 435–453.

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the time-based resource-sharing theory. *Psychonomic Bulletin & Review, 18*, 10–45.

Redick, T. S. (2019). The hype cycle of working memory training. *Current Directions in Psychological Science, 28*, 423–429.

Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology, 53*, 671–685.

Sala, G., & Gobet, F. (2019). Cognitive training does not enhance general cognition. *Trends in Cognitive Sciences, 23*, 9–20.

Schel, M. A., Thompson, A., & Steinbeis, N. (2020). *Development of behavioral and neural variability of cognitive control.* Manuscript submitted for publication.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Marrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest, 17*, 103–186.

Smithers, L. G., Sawyer, A. C. P., Chittleborough, C. R., Davies, N., Smith, G. D., & Lynch, J. (2018). A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nature Human Behaviour, 2*, 867–880.

Stamps, J. A., & Frankenhuis, W. E. (2016). Bayesian models of development. *Trends in Ecology & Evolution, 31*, 260–268.

Steinbeis, N. (2018a). Neurocognitive mechanisms of prosociality in childhood. *Current Opinion in Psychology, 20*, 30–34.

Steinbeis, N. (2018b). Taxing behavioral control diminishes sharing and costly punishment in childhood. *Developmental Science, 21*(1), Article e12492. doi:10.1111/desc.12492

Steinbeis, N. (2020). *Training cognitive control increases prosocial behavior.* Manuscript in preparation.

Steinbeis, N., & Over, H. (2017). Enhancing behavioral control increases sharing in children. *Journal of Experimental Child Psychology, 159*, 310–318.

Strobach, T., & Karbach, J. (Eds.) (2021). *Cognitive training: An overview of features and application* (2nd ed.). Cham, Switzerland: Springer.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). Cambridge, MA: MIT Press.

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review, 120*, 439–471.

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*, 106–113.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. III. Functions involving attention, observation and discrimination. *Psychological Review, 8*, 553–564.

Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research, 78*, 852–868.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33*, 67–81.

Wass, S., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory – Is younger, better? *Developmental Review, 32*, 360–387.