



HAL
open science

Computational mechanisms underlying the emergence of theory of mind in children

Thinh-Toan Axel Nong, Yansong Li, Chen Qu, Jae Hyung Woo, Yuqian Wang, Chengguo Miao, Xiaoyi Liu, Qingxia Liu, Jean-Baptiste van der Henst, Edmund Derrington, et al.

► **To cite this version:**

Thinh-Toan Axel Nong, Yansong Li, Chen Qu, Jae Hyung Woo, Yuqian Wang, et al.. Computational mechanisms underlying the emergence of theory of mind in children. 2023. hal-04288677

HAL Id: hal-04288677

<https://hal.science/hal-04288677>

Preprint submitted on 16 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Computational mechanisms underlying the emergence of theory of mind in children

Toan Nong^{1,2}, Yansong Li^{3,4,*}, Chen Qu^{6,*}, Jae Hyung Woo^{7,**}, Yuqian Wang^{3,**}, Chengguo Miao^{3,**}, Xiaoyi Liu^{3,**}, Qingxia Liu^{6,**}, Jean-Baptiste Van der Henst⁵, Edmund Derrington^{1,2}, Alireza Soltani⁷, Jean-Claude Dreher^{1,2}

- 1- *ISC-MJ, CNRS UMR5229, Neuroeconomics Lab, Bron, France*
- 2- *University of Lyon 1, UCBL1, Lyon, France*
- 3- *Reward, Competition and Social Neuroscience Lab, Department of Psychology, School of Social and Behavioral Sciences, Nanjing University, Nanjing, China*
- 4- *Institute for Brain Sciences, Nanjing University, Nanjing, China*
- 5- *CRNL-CNRS-UCBL1, Bron, France*
- 6- *Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, China*
- 7- *Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA.*
- 8- **, **: equal contributions*

Abstract (246/250 words)

Dyadic social interactions often require one to adapt to the fluctuating cooperative or competitive intentions of others, which can change over time without being explicitly signaled. This ability is a critical component of theory of mind (ToM), which develops gradually during childhood. However, the computations underlying this ability remain to be described at a mechanistic level. Here, we used a combination of experimental and modeling approaches in children (3 to 9 years old) to elucidate the computational mechanisms underlying this ability and to identify how these computations develop during childhood. We implemented a card-matching task in which 192 children (100 females) thought they were playing with another player. In fact, they played with an algorithm that alternated between cooperative and competitive strategies across blocks of trials without warning. Behavior of younger children (kindergarten, age 3-5) was more compatible with a simple decision process that tracked the probability that the other agent would choose a particular card, based on their previous actions. By 8 years of age, most children mainly followed a mixed-intentions influence learning (MIIL) model, that arbitrates between cooperative and competitive intentions to adapt their decisions. These results show that the combination of the simulation of one's own influence on others and of an arbitration process between cooperative and competitive strategies is an important developmental milestone occurring at about 8 years of age. Our findings characterize the development of the computations required to adapt to the fluctuating cooperative or competitive intentions of others during childhood.

Significance statement (114/120 words)

This study identifies the computations required for strategic interactions in children using a formal computational approach. We characterize the nature and milestones of the development of computational mechanisms involved in adapting to the intentions of others in children which is part of Theory-of-Mind. Most children become able to mentalize their influence on another's decision and adapt to another's intentions at about 8 years old. The model accounting best for behavior around this age was a model that arbitrated between a mixture of a competitive and a cooperative version of an influence learning model. This influence learning model is a reinforcement learning model that also integrates the influence of its decision on the other's outcome.

Introduction (1146 words)

Often considered as an essential requirement for sophisticated social interactions, Theory-of-Mind (ToM) is the ability to attribute mental states, such as emotions, beliefs, or desires, to other individuals¹. Different milestones have been proposed in the development of ToM from birth to early adulthood^{2,3}. Its most basic form, perception-goal psychology, emerges from around 9 months of age and allows agents to represent that others may have different perceptual perspectives on the world and act accordingly. Meta-representations emerge from around 4 years of age, in the form of belief-desire psychology, and involves an appreciation that others subjectively represent the world in fine-grained aspectual ways that may be incompatible with one's own view and that may be false⁴⁻⁷. At around 7 years of age emerges further key concepts that further refine children's meta-representational understanding, sometimes referred to as Advanced Theory-of-Mind (AToM)^{5,7-11}. It describes

multiple aspects of an advanced understanding of the minds of others, including social reasoning^{5,8}.

Here, we specifically focused on the ability to adapt to fluctuating cooperative versus competitive intentions of others, which may be an important cognitive component required for the social reasoning component of AToM. The ability to track how others' behavior might change over time, and to adapt to the hidden intentions of others that fluctuate between competitive and cooperative strategies, is essential to keep an advantage in social interactions. In such contexts, deciphering intentions not only requires the representation of beliefs about other's intentions, as in cooperative or competitive contexts alone, but it also requires deciding whether the other is competing or cooperating to deploy an appropriate behavioral response¹². Although formal computational models of ToM functions have been proposed in adults¹²⁻¹⁸, it is unclear whether these models apply to children and how they evolve with development during childhood from 3 to 9 years of age.

Characterizing the computational mechanisms underlying the ability to adapt to the fluctuating cooperative or competitive intentions of others is important because it may help build a mechanistic understanding of how this important component of ToM develops in childhood⁵. Different models of strategic interactions may account for the computational mechanisms engaged in this ability. To elucidate the computations underlying strategic decision making in children, we compared the predictions of different computational models. A simple and straightforward strategy may be to use simple heuristics to make decisions, such as the Win-Stay/Lose-Switch strategy, that consists of persisting with one's choices when winning, and switching choices after a loss¹⁹.

A popular theory is that the learning and decision processes are explained by a model-free reinforcement learning (RL) algorithm, such as a Q-learning model, that is able to learn the value of an action in a particular state. This algorithm allows to compute and update a Q-value that is derived from a prediction error that corresponds to the difference between the predicted value of the model and the actual feedback²⁰. A more sophisticated candidate model is the influence learning model^{17,18} that not only employs RL mechanisms to track outcome probabilities upon choice, but also accounts for a mentalizing process that computes the influence of one's choice on another's decision. A different class of algorithm relies on Bayesian inferences. Based on probabilistic representations of the world, it uses the results of its observations and priors on the environment to infer optimal actions^{21,22}. According to this latter class of model, children might learn and decide based on purely Bayesian learning strategies that do not rely on any mentalizing process. Finally, a recent promising computational model, known as the Mixed-Intention Influence Learning (MIIL) model¹², not only accounts for the influence of one's choice on another's decisions, but also ponders between competitive and cooperative opponent's decision strategies. The MIIL model outperforms other learning models in predicting social choice behavior in adults when they had to alternate between a cooperative and a competitive game. It relies on and arbitrates between a competitive and a cooperative expert to make strategic decisions, one assessing competitive intentions and the other assessing cooperative intentions. A controller then weights between these experts according to their relative reliabilities. This sophisticated neurocomputational mechanism may be absent or only partially developed in younger children.

In the current study, we sought to identify whether distinct computational mechanisms appear during children's development before they develop a deeper theory of mind. We therefore used a computational modeling approach and a strategic decision task similar to the one developed by Philippe et al in adults¹². Briefly, pairs of children completed a task in which they played a card-matching game on a tablet (**Fig.1**). Each child from these pairs played facing a same-gender child (**Fig.1.a & b**). Both were told that they were playing with each other, but were not allowed to communicate. Unbeknownst to them, they were in fact playing with an Artificial Agent (AA) that switched between cooperative and competitive trial blocks (two modes, see **Fig.1.d**). The goal of the participating children was to score as many points as possible in the game by guessing the correct color. The more points they scored the more reward they would obtain at the end of the experiment. This task allowed us to investigate the

computations used by children during social interactions and to identify at what developmental age they are able to adapt to fluctuating intentions of others while the modes of interaction (cooperation vs competition) were not indicated.

We first identified the computational processes that described the children's decision process in this dyadic social interaction task. Then, after fitting each of the candidate models to the collected data, we compared them to determine the best fitting model at the group level at different ages. We then analyzed the learning parameters of this winning model to characterize the evolution of its hidden states in our different age groups: Kindergarten (KG) from 3 to 5 years old, Early Primary Schoolers (EPS), 6-7 years old, and Mid-Primary Schoolers (MPS), 8-9 years old. This allowed us to describe at a mechanistic level how adaptation to changing intentions of others develop and change between different age groups.

At the behavioral level, we expected the ability to differentiate between the competitive vs cooperative blocks to increase with age. Children around the age of 8 should adapt to the hidden cooperative or competitive intentions of a partner based on the previous sequences of choices and outcomes. At the computational level, these expected results mean that younger age groups of children should be better fitted by simpler, less sophisticated models. We hypothesized that sophistication of the computational models describing the decision process should increase with age. According to this hypothesis, the most sophisticated model, i.e. the MIIL model, should appear in the oldest age group. Alternatively, if the same model fits all age groups best, the learning parameters of such model should lead to better performance and discrimination between competitive vs cooperative blocks in older children.

Methods

Participants

192 participants (aged 3.4 – 9 years old, $M = 6.5$, $SD = 1.5$, 100 females) were recruited from one primary school and one kindergarten in China, Guangzhou city and Nanjing city.

114 participants (aged 6-9, $M = 7.5$, $SD = 1.1$, 60 females; 58 aged 6-7, $M = 6.5$, $SD = 0.3$, 30 females; 56 aged 8-9, $M = 8.5$, $SD = 0.3$, 28 females) were recruited from one primary school in Nanjing, Jiangsu province, China. These participants had normal or corrected-to-normal vision and were naive with regard to the purpose of the experiment. They and their parents gave written informed consent prior to participation.

78 participants (aged 3.4-6, $M = 5.0$, $SD = 0.8$, 40 females) were recruited from a kindergarten in Guangzhou city, Guangdong province, China. These participants had normal or corrected-to-normal vision and were naive with regard to the purpose of the experiment. Their parents gave written informed consent prior to participation.

The study was approved by the Ethical Review Board of Nanjing and Guangzhou Universities.

Mixed intentions task

Participants performed a task comprising 154 trials on computer tablets (MS Surface Pro). They were led to believe that they were interacting with another child sitting in front of them, and who was playing the same game with them. In fact, both were playing against an artificial agent (AA) managed by a computer program. This simulated social interaction allowed us to investigate the dynamics and mechanisms arbitrating between multiple learning algorithms. The experimenter conveyed the hierarchy to the 2 children when they first presented them to each other. In addition to their names, the experimenter also clearly stated the school grade of each of them. Males played with males, and females with females. The tablet screens showed each participant four cards, two face-down (the other player's cards) and two face-up (their own cards). Participants were informed that to win, they had to choose the card of the same color as the one the other person was going to choose. Experimenters were careful not to specify whether the other player was an opponent or a partner i.e. whether the other player had the same victory condition (to match cards) or whether they would win if the cards did not match. Participants were also told that both they and the other player had to make their choices in two seconds (**Fig.1**). If the Artificial Agent (AA) played before the participant, one of the two face-down cards was removed from the playing field. If the participant chose first, only the selected card remained on the playing field. Then, when both had chosen, the cards were revealed. The participant scored points if the card colors matched, otherwise they received nothing. Participants were led to believe that their final payoff would be increased if they scored more points. No information about the other's payoff was given to the participants, they only knew that after an interaction, the other 'participant' would see the same screen but with their own outcomes which could be different for each of them.

Importantly, unbeknownst to the participants, the artificial agent alternated between Competitive and Cooperative trial blocks. During this mixed-intentions task, the AA's strategy was determined by alternating 20 trials of a coordination game (Cooperative blocks), and then 10 trials of a hide and seek (HS) task (Competitive blocks). The artificial agent algorithm was designed to predict the color that would be chosen by the participant on the basis of a probabilistic analysis of the two previous choices and outcomes (see **Supplementary Methods** for the algorithm). Here we defined a competitive choice, made by the AA, as choosing the card of the color the participant was expected not to play, and a cooperative

choice as choosing the card with the same color that the participant was expected to play. Thus, the artificial agent exploited the bias of the participants in stochastically, such that the more predictable the participant was, the more the algorithm made correct competitive or cooperative choices (see **Supplementary Methods**). Participants were not informed of the switches between cooperative and competitive interaction by the algorithm.

Artificial Agent

The AA calculated the probability p for the participant to select a particular target color based on the history of the two previous choices and their outcomes. It then chose the target that it deemed the most probable to be chosen by the participant in the cooperative mode, and the target that was the least probable in the competitive mode.

Behavioral analysis

Logistic regression: For the logistic regressions, we reported significant marginal effects of a given variable under the name “estimate” (for example: Cooperativity signature_{t-1} : estimate).

$$\ln\left(\frac{p}{1-p}\right) = x_0 + x_1X_1 + x_2X_2 + \dots$$

X_i represents independent variables and x_i the associated weights in the logistic regression. P represents the probability of a given event. The marginal effect of the variable X_1 is defined as:

$$\widehat{y}_1 = \text{mean}(\text{logit}^{-1}(x_1))$$

The mean is computed across all observed data. Thus, the marginal effect called “estimate” can easily be interpreted as the discrete change of the dependent variable given a unitary change of an independent variable. For the linear regressions, reported “estimate” represents x_i , i.e., the regression coefficient. Indeed, in a linear regression, the marginal effect of a variable is equal to the estimated coefficient.

Information theoretic metrics: To quantify the consistency in the adopted strategy in response to reward feedback, we utilized our previously developed metrics based on information theory^{23,24}. These include entropy of choice strategy ($H(\text{Str})$), mutual information between reward outcome and strategy (MIRS), and conditional entropy of reward-dependent strategy (ERDS). Intuitively, the entropy of choice strategy ($H(\text{Str})$) captures the randomness or uncertainty in the adopted strategy, trial-by-trial. Thus, it signals the overall stochasticity in the choice behavior. MIRS and ERDS values further aim to constrain the extent to which the agent’s choice strategy can be predicted by the reward outcome prior to that decision. Importantly, in our dyadic game settings, the reward feedback is directly predicated upon the other player’s action and therefore could signal additional social value.

Specifically, the Shannon entropy of choice strategy, $H(\text{Str})$, is computed using the following equation:

$$H(\text{Str}) = -(P(\text{stay}) \cdot \log_2 P(\text{stay}) + P(\text{switch}) \cdot \log_2 P(\text{switch})),$$

where Str is the agent's adopted strategy, coded as *stay* (1), if the agent selects the same target as the previous trial, or *switch* (0) otherwise. Therefore $H(Str)$ measures the level of uncertainty or surprise choice behavior in terms of stay or switch.

Mutual information between reward and strategy (MIRS) is calculated as follows:

$$MIRS = I(Str; Rew) \\ = - \sum_{Str \in \{stay, switch\}} \sum_{Rew \in \{win, lose\}} \left(P(Str, Rew) \cdot \log_2 \left(\frac{P(Str, Rew)}{P(Str) \cdot P(Rew)} \right) \right),$$

where Rew is the reward outcome on the previous trial, coded as win (1) or lose (0).

Finally, the entropy of reward-dependent strategy (ERDS), is the remaining uncertainty in the strategy after accounting for the information given by previous reward outcome:

$$ERDS = H(Str|Rew) = H(Str) - I(Str; Rew) = \\ - \sum_{Str \in \{stay, switch\}} \sum_{Rew \in \{win, lose\}} \left(P(Str, Rew) \cdot \log_2 \left(\frac{P(Str, Rew)}{P(Rew)} \right) \right).$$

These metrics are model-free in the sense that they do not assume any underlying structure or model in the learning and decision-making processes. As such, these measures can be computed directly from any segment of the task, making them especially useful for quantifying behavioral adjustment over time in dynamic, changing environments. For the running average plot, we computed the metrics for each individual using a moving window of 7 trials, such that the data point for trial t was computed from trials between $t-6$ and t . The values were then averaged for each age group. For calculating the bar plots in the inset, we calculated the metrics for each phase in each subject by compiling the last 10 trials of each Cooperation block and the last 5 trials of each Competition block (i.e., last half of each block) to reduce the contaminating effect of the preceding phase. These values were then averaged across subjects by age groups.

Learning models

To test for a dynamic tracking of implicit intention we compared 5 models, 1 involved theory of mind (the MIIL model), the others were to control for other possible strategies (see Supplementary Materials). The influence model relies on Taylor expanded reinforcement learning^{17,18} to take into account the influence of one's own strategy on the strategy of the other. This model can be adapted in a cooperative or in a competitive version. The MIIL model is constructed based on an arbitrator that weights the cooperative and competitive versions of the influence model according to their reliability before making the decision¹².

To control for strategies that did not include the ability to adapt to changing intentions of others (theory of mind), and thus could not differentiate between cooperation and competition, we added 4 other models including two Bayesian inference types (BSL). The Bayesian Sequences Learner (BSL) strategy relies on Bayesian inference given past sequences of choices. We used sequences of depths 0 and 1. Finally, we added two non-Mentalizing non-Bayesian models, a reinforcement learning model (QL) and a model based on the heuristic Win/stay – Lose/Switch that we observed in the model free analysis.

The Bayesian Model Selection (BMS) was performed using the VBA toolbox (Variational Bayesian Analysis) in a random effect analysis relying on the free energy as the lower bound of model evidence. We use Exceedance Probability measurements (EP)²⁵ to select the model used most frequently in our population.

Results (2162 Words)

Behavioral signature of tracking intentions.

To study possible signature of tracking intentions, we considered multiple variables that could affect choice behavior. These include the previous stay vs switch choices of the child and of the artificial agent (AA) in the previous three trials (at $t-1$, $t-2$, $t-3$), the three previous outcomes (i.e. whether the previous three trials were won or lost), the age and sex of the child, decision time, the hierarchical condition (i.e., if they were paired with a child in the same school grade, or an older child, or a younger child), the number of trials within a session of the experiment, the strategy of the AA at that time (competitive or cooperative), the previous cooperativity signatures at $t-1$ and $t-2$ (see below), and finally the interactions between age and cooperativity signatures as regressors. We first investigated the probability to “stay”, i.e., to select the same target as that from the previous trial, for the three different age groups (**Fig. 2a**). We examined how children used the history of previous interactions to make their choices. To measure cooperation, we defined the “cooperativity signature” as a binary variable that is equal to 1 if:

- the participant won on the previous trial and the artificial agent (AA) stayed on the same target for the next trial, or,
- the participant lost on the previous trial and the AA switched to the other target for the next trial.

Otherwise, the cooperativity signature was set to 0. Indeed, from the perspective of the participant, if the AA is a cooperative partner, then when both players win by choosing the same target, they should choose to keep the same target. Instead, if they lose due to choosing different targets, the AA’s switch can be seen as its willingness to conform to the participant’s choice. Cooperativity signature equal to 1 corresponds to the AA following win-stay/lose-switch strategy.

We first computed the mean behavioral probability to “stay”, i.e., to select the same target as that from the previous trial, for the three different age groups (**Fig. 2a**). To control for random effects and the interaction and autocorrelation between effects of each variable over time, we used Generalized Linear Mixed-Effects (GLME) models implemented with the lme4 package in R. We ensured that we had good estimates of random effects and accounted for variability in behavior using Bound Optimization by Quadratic Approximation (see **Methods**). GLME1 assessed the children’s tendency to select the same target as that in the previous trial (“stay”) and GLME2 assessed their performance in successfully finding the same card as the AA (“reward”).

When examining how children used the history of previous interactions to make their choices, we found that the results of previous three trials, i.e., the decisions of both the participant and the AA to stay and whether the participant was successful, significantly predicted the probability that the participant would stay and win, independently of all other factors (**Supp.Tab.1**). Participants’ performance improved with time, and, within a session of the experiment. Children tended to stay more frequently over time in the cooperative blocks. This suggests that their performance improved with practice. Interestingly, the cooperativity signature at $t-1$ predicted a higher probability for the children to stay with their choice. This indicates that children tracked whether the other agent was cooperating during the previous trial. More specifically, the marginal effects of the interaction between cooperativity signature and age (**Fig. 2b left**) indicated that older children performed significantly better after a positive (vs zero) cooperativity signature from the AA. In comparison, after controlling for other independent variables, younger children did not show such improvements in performance after trials with a positive cooperativity signature. This indicates that older children identified

signatures of cooperativity better. In contrast, younger children also tended to stay less in the cooperative blocks, independently from other factors. However, taking into account the marginal effects of the interaction between the age of the child and the mode of the AA (cooperative vs. competitive) (**Fig. 2b right**), we observed that stay behavior was more frequent in cooperative blocks compared to the competitive blocks for older children, whereas, surprisingly, it decreased in cooperative blocks compared to competitive blocks for younger children. These results point towards an evolution of the learning process for intention adaptation and coordination strategy. We found no effect for sex or hierarchical condition. These results persisted after a step-by-step drop-term procedure or addition of a random effect of time.

Model-free analysis based on information theory reveals emergence of new strategies in older children

To examine whether and how children continuously adjust their behavior across cooperative and competitive blocks and reveal different choice strategies across age groups, we next computed several metrics based on information theory to measure consistency in children's responses to the AA's choice and reward feedback. More specifically, we computed the entropy in participant's choice strategy in terms of stay or switch on the previous target ($H(\text{Str})$), mutual information in reward-dependent strategy (MIRS), and conditional entropy of reward-dependent strategy (ERDS) over the time course of the experiment. Children were grouped as kindergarteners (KG), early primary schoolers (EPS), and mid-primary schoolers (MPS). We computed these quantities using a running average (with window size of seven trials) and separately for each age group (**Fig. 3**, see Methods for more details).

We found that during the cooperative blocks, better performance (**Fig. 3a**) in older children was accompanied by lower $H(\text{Str})$ (Spearman's correlation; $r = -0.718$, $p = 7.15e-32$), suggesting that older children were overall more consistent in their stay or switch between trials (**Fig. 3b**, Wilcoxon rank sum test on $H(\text{Str})$; KG vs. EPS: $p = .0169$; EPS vs. MPS: $p = .0203$). The lower $H(\text{str})$ observed in older children was mostly due to a decrease in ERDS ($r = 0.852$, $p = 1.53e-55$; **Fig. 3d**), corresponding to more consistent responses to reward feedback, and to a lesser extent, a decrease in MIRS ($r = 0.438$, $p = 1.98e-10$; **Fig. 3c**) corresponding to a weaker link between choice and previous reward outcome. These results indicate that the superior performance of older children during the cooperative blocks was due to two factors: (1) older children were able to choose more consistently across trials (less stochasticity in choice); (2) older children based their choices on reward outcome less strongly.

In contrast, the differences between the three age groups were smaller during the competitive blocks. More specifically, there were no significant differences in their consistency in strategy ($H(\text{Str})$), even between the youngest (KG) and the oldest (MPS) age groups (Wilcoxon rank sum test, $p = .582$; **Fig. 3b**, inset). Yet, we observed that the oldest (MPS) age group responded more strongly to the outcome on the previous trials, as indicated by significantly larger MIRS than the younger EPS group (Wilcoxon rank sum test, $p = .0285$; **Fig. 3c** inset) or KG group ($p = .038$). Similarly, the randomness in strategy after accounting for the effect of reward outcome (ERDS) was significantly smaller for the oldest group (Wilcoxon rank sum test, $p = .00205$; **Fig. 3d** inset), but not significantly different between KG and EPS children ($p = .311$). The smaller ERDS for the oldest group compared to the younger group (EPS) was accompanied by a larger MIRS (**Fig. 3c**). These data indicate that during the competitive blocks, only children in the oldest group were able to learn better from reward feedback and this made them more consistent in response to rewards.

Overall, these results suggest that only the oldest group of children (MPS) were able to adjust their strategy appropriately when the intention of AA changed between the cooperative and competitive blocks. Consistent with this interpretation, we observed that the MPS group showed the largest overall increase in MIRS from cooperative to competitive block (**Figure 3c**; mean increase = 0.0788; Wilcoxon signed-rank test, $p = .00127$), whereas the younger age groups exhibited insignificant changes in MIRS (KG: mean increase = -0.00286, $p = .891$; EPS: mean increase = -0.0438, $p = .634$). These results indicate the development of additional strategies in older children that allow them to learn from reward feedback differently depending on the nature of the interaction with the other agents.

Computational models track children's evaluation of the "other player's" intentions

To elucidate the computations underlying the children's strategic decision-making, we compared a variety of different computational models fitted to the choice data of individual children on a trial-by-trial basis. We sought to examine their ability not only to mentalize the other agent, but also to adapt to its intentions. Therefore, we fitted different Influence models^{17,18}, including the Mixed-Intentions Influence Learning (MIIL) model¹². Classical influence models rely on mentalizing mechanisms that account for the influence of one's own choice on the other's choice. We tested a competitive as well as a cooperative version of the classical influence model. In contrast, the Mixed-Intentions Influence model computes one decision value according to a competitive expert and another according to a cooperative expert. It then arbitrates between the two, based on the difference in their relative reliability, defined as the difference in unsigned value functions for the two choices determined by specific learning algorithms (see Supplementary Material). We compared these models to other models that did not rely on mentalizing mechanisms (see Methods). Specifically, we fitted a Win-Stay/Lose-Switch (WS/LS) heuristic¹⁹, a Q-Learning (QL) model based on reinforcement learning, and Bayesian models of different complexity, (Bayesian Sequence Learners of depth 0 and 1 (BSL0 and BSL1)²⁶ (see **Supplementary Methods** for more details on the models). All of these models were initially fitted separately to each participant and in each condition. We applied a Variational Bayes Approximation to fit and optimize all of the tested models based on the Free Energy criterion²⁷ with the VBA-toolbox²⁸ in Matlab. This ensured we had good fits of the tested models to the data and accounted for variability in behavior (see **Supplementary Methods**). We then compared all the tested models with a group-level random-effect Bayesian Model Selection (BMS)^{25,29} based on the Free Energy criterion with the VBA toolbox.

We found that the best fitting model for the majority of children in groups KG and EPS was the BSL0 model (**Fig.4a&b**), which is a Bayesian learning model that computes the mean and variance of the probability of the correct choice for the next trial (KG: Estimated Frequency $E_f=0.605$, protected Exceedance Probability $p_{EP}=0.999$; EPS: $E_f=0.530$, $p_{EP}=0.986$). These results show that a majority of the children from 3 to 7 years old neither tracked the intentions of the other agent, nor relied on any mentalizing process for their decision in this task. They rather tracked with a simple Bayesian strategy which card was the most probable at each trial. In mid-primary schoolers (MPS), the MIIL model fitted the majority of the children best (MPS: $E_f=0.483$, $p_{EP}=0.956$, **Fig.4a&b**). This indicates that the ability to mentalize and to track the intentions of others for strategic decision-making actually appears around the age of 8 years old, when children are in the third year of primary school (in China). Furthermore, this ability to mentalize and track intentions, with an un-signaled alternation between cooperative and competitive blocks, appears gradually among children starting from kindergarten (KG: $E_f=0.122$, $p_{EP}=10e-11$; EPS: $E_f=0.278$, $p_{EP}=1.36e-2$; MPS: $E_f=0.483$, $p_{EP}=0.956$, **Fig.4a&b**). According to these findings, children progressively develop the ability to mentalize and update their beliefs about future chosen targets, and arbitrate between the predicted

intentions of another to compete or cooperate. This ability appears earlier in some children (at kindergarten) but the majority of mid-primary schoolers apply this decision process which remains the most common strategy in adults ¹².

The sophistication of children's decision process increases with age

The BSL0 model best accounted for the behavior of the majority of younger children, but was almost entirely absent in the oldest group. Thus, the model frequencies of the oldest (MPS) vs youngest (KG) groups of children differed significantly. The posterior probability that the two groups had the same model frequencies was $p = 6.7e-3$ (uncorrected). Although there was no significant difference between the frequency of the different models in the KG and EPS groups ($p=0.995$) or in the EPS and MPS groups ($p=0.990$), there was a general trend towards an increase in the sophistication of the strategies such that the BSL1 model best accounted for the behavior of more EPS and MPS students than the BSL0. The WSLS heuristic remained a marginal strategy for some EPS and MPS participants, although its representation increased among the other models with age (see **Fig.4a&b**).

Finally, analysis of the arbitration parameters of the MIIL model for the children that were best fitted by this model, revealed how the model attributes the cooperative vs competitive tendency of the children (**Fig.4c**). In general, the children showed a low tendency to cooperate in the game. This was the case across the three age groups. Their tendency to cooperate increased during the cooperative blocks, and interestingly, in the older the age group, children became more likely to cooperate than not during the cooperative blocks. Furthermore, for the MPS group, the tendency to cooperate in cooperative blocks increased progressively over consecutive cooperative blocks. This indicates that the older children improved their ability to decipher the cooperative intentions of the artificial agent in the cooperative blocks during the task. Thus, our results indicate differences in both model frequency between the age groups, but also that the children best fitted by the MIIL model were also more capable of recognizing cooperative intentions.

Discussion (1356 words)

Learning to adapt one's own decisions to those of another facilitates social interactions. The abilities to adapt to the intentions of others, to determine the influence of one's actions on the actions of others, and to apply strategic reasoning during social interactions are relevant AToM abilities for children's everyday experiences. Such skills are required for all sorts of competitive or cooperative interactions in the playground and classroom at school. Here, we first examined how children across age groups adjust their choice behavior across cooperative and competitive blocks. Next, we characterize the computations required to adapt to changing intentions of others in children. We demonstrate that by the age of 8 years old (i.e. mid-primary school), most children used a computational mechanism relying on a mixed-intentions influence learning model, that arbitrates between cooperative and competitive intentions to adapt their decisions. The strength of this approach was to compare diverse computational models to assess how children learn to adapt their decisions to those of another. This computational approach accurately captures the development of AToM abilities during childhood at a mechanistic level, avoiding issues of interpretation that can be encountered when relying only on behavior^{10,11,30}.

We found that when a task is not explicitly signaled as cooperative or competitive, most children in mid-primary school exploit a cooperativity signature, i.e. intention to cooperate from their partner, better than younger children. Measures of entropy and mutual information based on information theory indicate that older (vs younger) children performed especially well in cooperative (vs competitive) blocks due to 1) their higher choice consistency and 2) their weaker reliance on the previous outcome to decide. Older children thus discriminate better between competitive vs cooperative blocks by behaving more predictably in the cooperative blocks than in the competitive ones, and by displaying more forgiveness in these blocks. At the computational level, we furthermore provide evidence that most children in mid-primary school, are able to track variations of another individual's hidden cooperative/competitive intentions. Their behavior was best explained by a Mixed-Intentions Inference Learning (MIIL) model¹² which exploits a control mechanism that gives a higher weight to the strategy with the most reliable prediction³¹. In contrast, the best fitting model for the majority of children in KG and EPS groups (children from 3 to 7 years old) was a learning model that computes the mean and variance of the probability of the correct choice for the next trial. These age-group of children neither tracked the intentions of the other agent, nor relied on any mentalizing process for their decision in this task. They rather tracked with a simple strategy which card was the most probable at each trial.

Computational modeling helped to reveal the mechanisms underlying learning to adapt to the changing intentions of the other. The intention of the children to cooperate or compete was captured by the arbitration weight factor of the MIIL model. This factor evolved with time depending on the competitive vs cooperative blocks of the AA for most children in the MPS group. More precisely, the intention to cooperate during cooperative blocks grows stronger with time over the course of consecutive blocks. This suggests that children understanding of the intentions of the other player, in this case the AA, improved with time during the task, and that this was mainly driven by the realization that the other player could cooperate or compete. This indicates that these children can actively learn and adapt to the intentions of the other player. Very few of the children in the younger groups showed this ability during the task, although it was apparent in some children in the intermediate EPS group. However, their ability to learn and adapt to the AA, as captured by the arbitration weight factor, was far lower than for the children in the oldest group. When considering the arbitration weight of the MIIL model (**Fig.4.c**), the oldest group of children differentiated between competition vs cooperation

intention of the AA better than the two younger groups. Finally, it should be noted that in all three age groups there was a strong bias towards competition. This might be because most videogames played by children at this age, on a tablet and also in their daily experience, involve competition in one way or another. The competitive bias might also have been fostered by the experimental setup in which children were placed facing each other, which might encourage a competitive, rather than a neutral context, for children.

Our computational modeling analysis indicates a shift of nature of the algorithms (from BSL0 to MIIL) used by children between early primary school (EPS) and mid-primary school (MPS). This is consistent with behavioral reports that core aspects of AToM develop nonlinearly, with children reaching a milestone at the age of 7 years, at which time they attain the conceptual insight that mental states can be recursive⁵. Indeed, it has been proposed that a distinct conceptual development occurring in middle childhood at around 7-8 years old could explain improved performance involving social reasoning and reasoning about ambiguity. In contrast, a simple general increase in information-processing capacities would have only predicted a change from BSL0 to BSL1 during development, and cannot explain such conceptual development in reasoning. Our findings therefore confirm that a conceptual development occurs in children between early primary school (EPS) and mid-primary school (MPS) age. We were able to characterize this change at the computational level by showing that it is underlined by algorithms of distinct nature. A conceptual insight might also explain the improved performance in our task, as most children in the oldest group have achieved the insight that: a) their own previous choices can influence others behavior and, b) that others can change their intentions during interaction and consequently switch strategies. We find that this conceptual insight occurs at the same age, 8 years old, as in previous large-scale studies⁸. This is an important period of cognitive development that might be linked to other complex forms of reasoning, such as scientific reasoning^{6,32,33}.

The MIIL model that fits best most MPS children is also the model that fitted best most of adults that performed the same task¹². Children above 8 years old are, in this sense, more similar to adults than younger children. This indicates that most 8-year-old children have developed the computational mechanisms that they will most likely maintain in adulthood to solve similar decision tasks. A similar trend was observed for other repeated economic games (stickers game and sender-receiver game) where children older than 6.5 years of age become more similar to adults than children younger than 6.5 years old³⁴. Our results would therefore indicate that most 8-year-old children have developed the computational mechanisms that they will most likely maintain in adulthood to adapt to others' fluctuating intentions or to solve similar decision tasks.

To conclude, our work provides evidence that important conceptual developments occur in children at around 8 years old and that they can be accounted by the Mixed-Intentions Inference Learning model. We established that, from mid-primary school onward, most children develop the abilities to adapt to the fluctuating intentions of others to cooperate or compete, and to learn to adapt one's own strategy as a consequence. They become able to mentalize the influence of their own choices on other's decisions and to arbitrate between a cooperative and a competitive strategy in a decision game that encourages adaptation to the hidden intentions of others. We document the computational mechanisms and the period when specific aspects of AToM for social reasoning develop during childhood, with a conceptual change in the decision-learning process between the youngest group and the oldest group of children. This suggests that conceptual developments fuel some aspects of AToM development in children^{4,5,8}. Research studies have observed diverse consequences of AToM, e.g., children's everyday social experiences³⁵, their self-understanding^{36,37}, as well as their epistemological beliefs³⁸ or their academic abilities^{39,40}. These abilities are fundamental for

children's psychological well-being and their success in life. Therefore, understanding the computational mechanisms underlying components of the AToM during childhood has practical relevance to optimize educational and social environments, especially for e-learning applications modeling children as computational agents, and may also help identify children at risk.

Acknowledgements

This work was supported by the Fondation de France (n°00089590) to JC D, by the National Science and Technology Innovation 2030 Major Program to Y.L. (No 2022ZD0205100) by a grant from National Natural Science Foundation of China to CQ (31970982; 32171019), and by a US National Institutes of Health grant to A.S. (R01 DA047870). This research has also benefited from the financial support of IDEXLYON from Université de Lyon (project INDEPTH) within the Programme Investissements d'Avenir (ANR-16-IDEX-0005) and of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency.

Author contributions

J.-C.D. and T.N. developed the general concept, experiment and models. Y.W., C.M., X.L., Q.L. performed the experiments under the supervision of Y.L. and Q.C. T.N. programmed the task under the supervision of J.-C.D. T.N. developed the models and implemented the algorithms, and analyzed the data under the supervision of J.-C.D. J.H.W. helped to analyze the data under the supervision of A.S. T.N. wrote the draft. E.D., J.B., J.H.W., J.-B.V.D.H., Y.L, A.S., and J.-C. D. edited the manuscript.

Competing interests

The authors declare that they have no competing interests.

Resources availability

Lead contact

Further information and requests for resources and inquiries should be directed to and will be fulfilled by the lead contact, Jean-Claude DREHER (dreher@isc.cnrs.fr). This study did not generate new unique reagents.

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data and codes supporting the main results are available at:

https://osf.io/nf2ma/?view_only=f90dac33122044d19d7e7dea1252c0a2

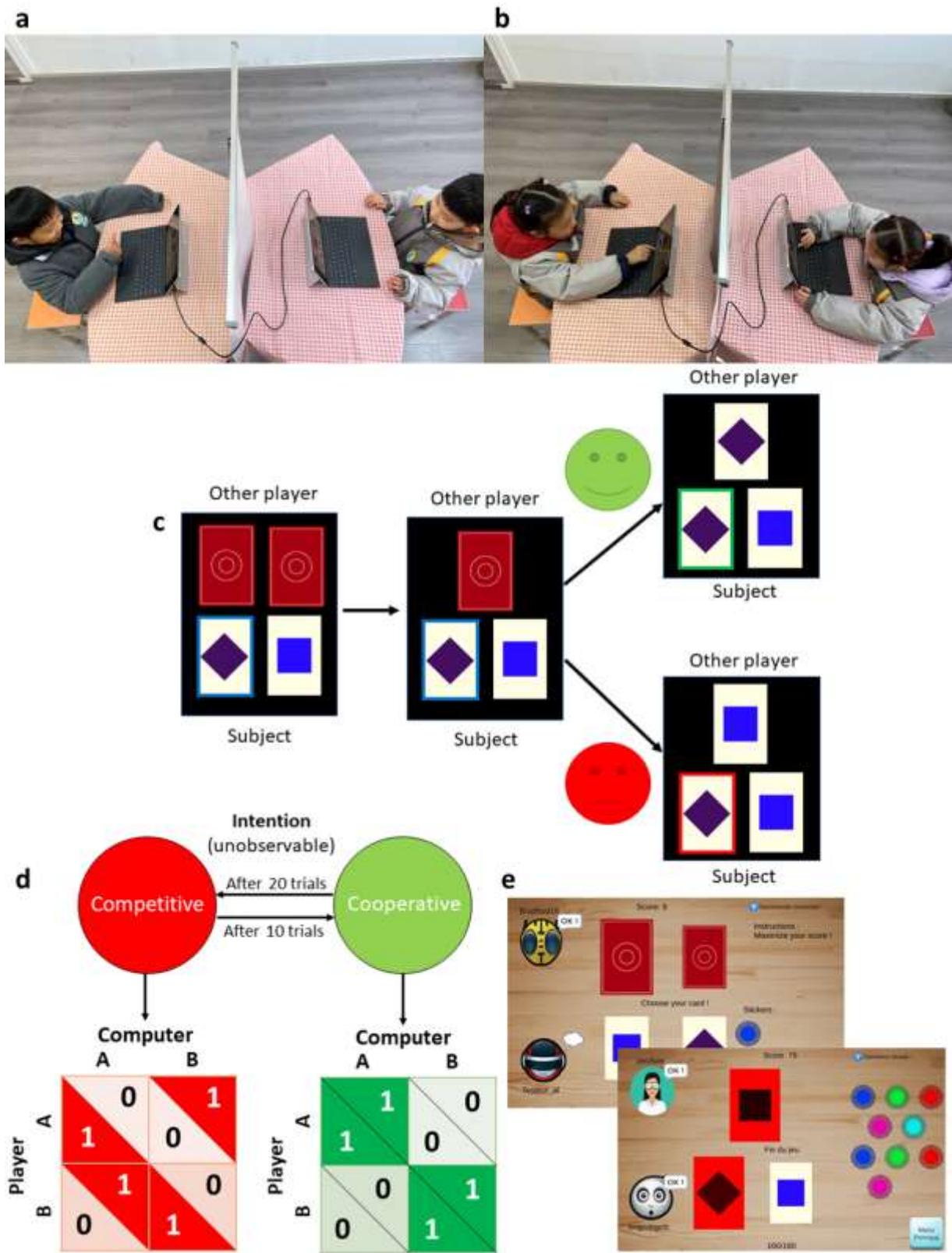


Fig. 1. Description of the procedure to study adaptation to fluctuating competitive and cooperative intentions of others in children. (a)-(b) Experimental setup. Two children were tested on individual tablets. They played facing each other, so each could see the other, but not the other's tablet. They were told that their tablets were connected, that the other player represented on their tablet was the child sitting opposite, and that each of them had to maximize their own score to maximize their final reward. In reality both children were actually

playing with an Artificial Agent (AA). Males were paired with males and females with females. **(c) Experimental protocol.** The two players both chose one card among two cards presented. When both children had chosen, the chosen cards were revealed, and if the card they had chosen was the same as that chosen by the AA, they scored points. If their card was not the same as that indicated for them by the AA, that child was not rewarded. **(d) Payoff matrix for the two possible modes of the Artificial Agent (AA).** The AA that decided which choice was rewarded for the participant operated in either a competitive or a cooperative mode, based on the payoff matrices depicted above. **(e) Snapshots of the actual game on the tablets.** The player chose one of the two cards presented on the lower part of the screen. The jewels on the right represent the number of rewards (images) that the child will be allowed to choose and keep at the end of the experiment.

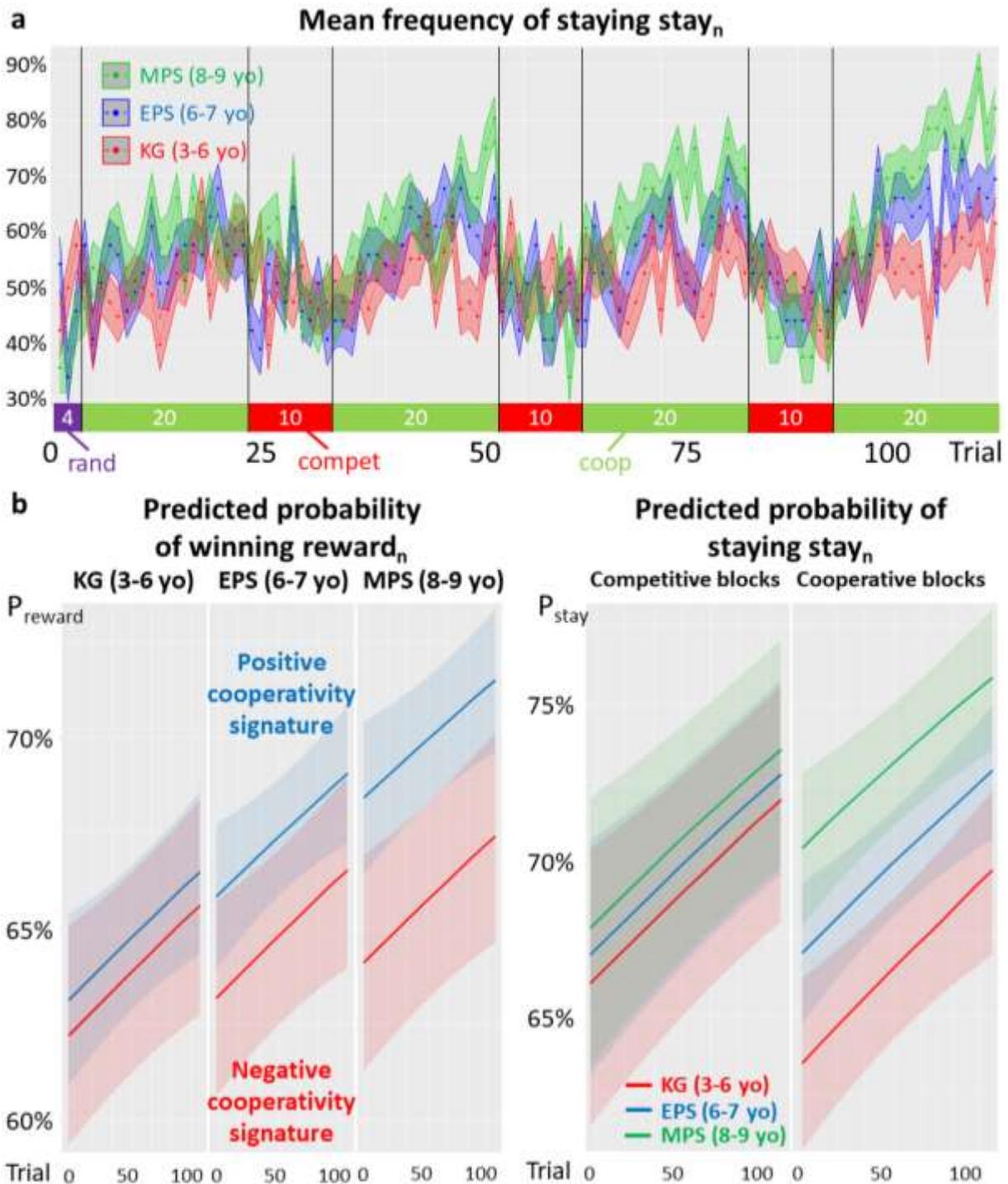


Fig. 2. Behavioral results. (a) Mean probability for the children to choose the same target as the one they chose in the previous trial (“stay”). The mean frequency to “stay” is represented in green, blue and red for children in mid-primary school (MPS), early primary school (EPS), and kindergarten (KG) respectively, with their respective standard errors (ribbons). The purple bar represents the first 4 initial trials in which the Artificial Agent (AA) played randomly for initialization purposes. The green bars represent the cooperative blocks and the red bars the competitive blocks. **(b) Generalized Mixed Effects (GLME) models, marginal effects. Left:** predicted probability for the subjects to choose the correct target at trial n , reward _{n} , plotted against the trial number (x-axis), after a positive cooperativity signature (blue) vs after a negative cooperativity signature (red) for different ages. **Right:** predicted probability for the subjects to choose the same target at trial n as previously (at trial $n-1$), stay _{n}

(right). The older the child, the better they succeed in coordinating with the AA after a positive cooperativity signature and the more stable they behave in the cooperative blocks.

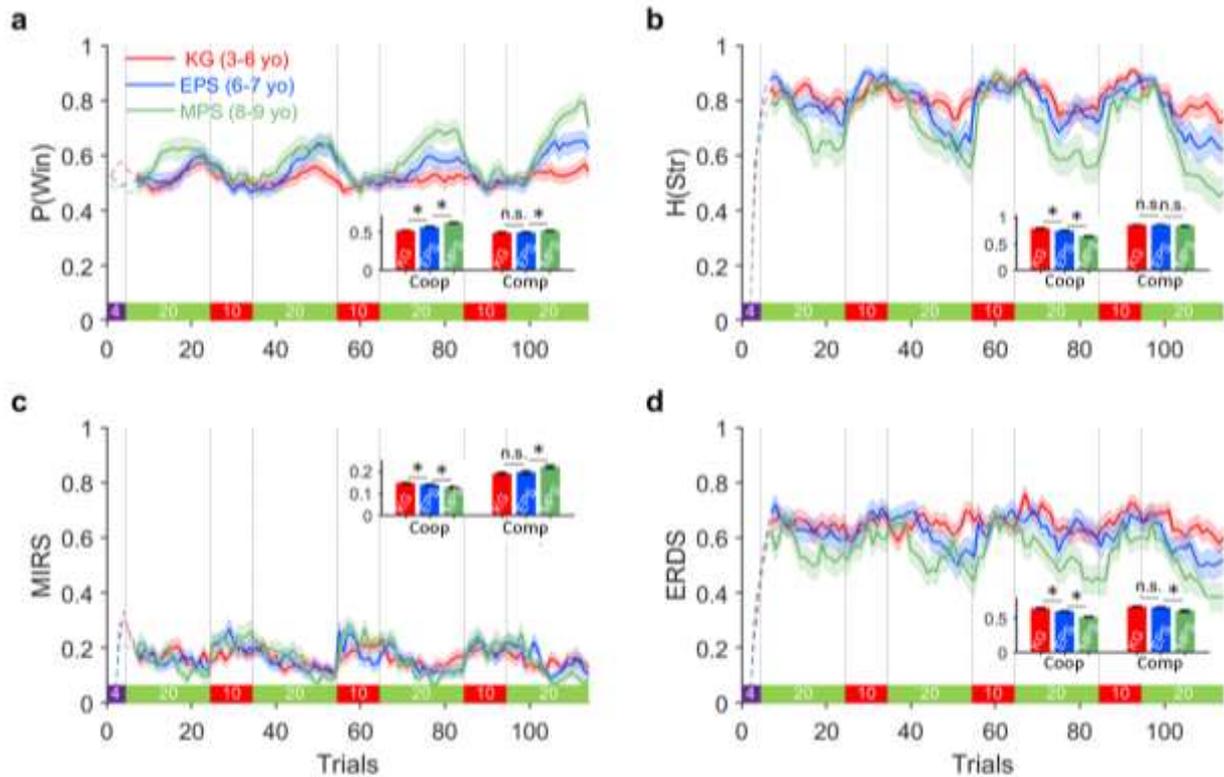


Fig. 3. Model-free analysis based on information theory. (a) Probability of winning reward over trials. Running averages of the probability of reward are computed using a moving window of 7 trials. Performance steadily increased with age during the cooperative blocks, but only slightly increased in older children in the competitive blocks. Inset shows the mean value for each group during each phase (Cooperative and Competitive). **(b) Shannon entropy for stay vs. switch over trials $H(Str)$.** During the cooperative phase, improvement in performance with age was accompanied by decrease in $H(str)$ as older children were better at choosing less randomly. No overall effect was observed during the competitive blocks. **(c–d) Mutual Information between reward outcome on the previous trial and the adopted choice strategy (MIRS) and entropy of reward-dependent strategy (ERDS)²³.** During the cooperative blocks, both MIRS and ERDS decreased in older children, showing that older children chose less randomly, and that they were also able to base their choice strategy on reward outcome less strongly. In contrast, during the competitive blocks, larger MIRS in 8-9 yo compared with 6-7 yo was accompanied by smaller ERDS in 8-9 yo compared with 6-7 yo. (Shaded error bar = S.E.M.)

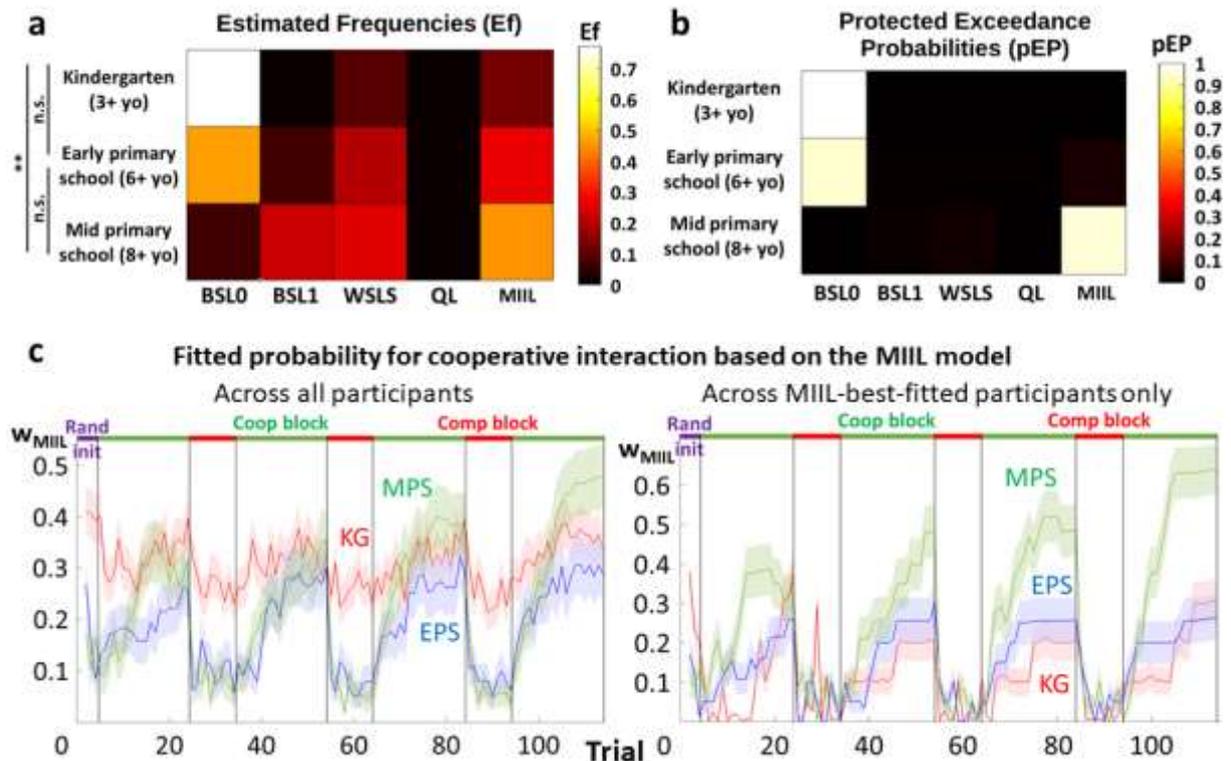


Fig. 4. Computational learning models. (a)-(b) Bayesian model comparison. x-axis: model names; y-axis: age group; Color-scale: (a) estimated frequencies (Ef), i.e., the frequency at which one model fitted one subject best in the group population compared to the other tested models; (b) protected exceedance probabilities (pEP), i.e., the probability that measures how likely it is that any given model is more frequent than all other models in the comparison set, corrected for the possibility that observed differences in model evidences (over subjects) are due to chance²⁵. In the Kindergarten (KG) and Early Primary School (EPS) groups, the model that best fitted most of our group of participants is the BSL0. In Mid-Primary School, the MIIL model best fitted most of the participants. **(c) Mean estimated cooperation probability computed from the MIIL model.** x-axis: trial number. y-axis: w_{MIIL} , the mean probability, according to the MIIL model, that the participant attempts to cooperate (left) across all participants (group by group) and (right) across only those participants that were best fitted by the MIIL model for the 114 trials (group by group). This indicates that children in Kindergarten fail to use sophisticated MIIL computations. The initial purple area shows the 4 random initializing trials, green areas are the Cooperative blocks, and red areas are the Competitive blocks.

Supplementary Materials

Learning Models

Bayesian Sequence Learning (BSL) model

The BSL model²⁶ tracks the (log) odds of $P(o_t=1|o_{t-K})$, where o is the partner's choice (binary outcome). This variable is updated according to a Laplace-Kalman filter, yielding 2 sufficient statistics (mean and variance) per combination of past outcomes. BSL can learn sequences of arbitrary depth K . For example:

- if $K=1$, then BSL tracks 2 probabilities, namely: $P(u_t=1|u_{t-1}=1)$ and $P(u_t=1|u_{t-1}=0)$. In this case, BSL needs to know about the partner's previous move u_{t-1} .
- if $K=2$, then BSL tracks 4 probabilities, namely: $P(u_t=1|u_{t-1}=1, u_{t-2}=1)$, $P(u_t=1|u_{t-1}=0, u_{t-2}=1)$, $P(u_t=1|u_{t-1}=1, u_{t-2}=0)$ and $P(u_t=1|u_{t-1}=0, u_{t-2}=0)$. In this case, BSL needs to know about the partner's two previous moves u_{t-1} and u_{t-2} . etc.

More generally, the BSL model tracks 2^K probabilities.

Win-Stay Lose-Switch (WSLS) model

This heuristic¹⁹ keeps the same target as the previous one if previous reward was positive, and changes targets otherwise.

Q-Learning (QL) model

The constituents of this model in its simplest form are:

- a set of (action/item) hidden value states. In two-armed bandit problems, there are two of these (n in a n -armed). In general, there will be as many values as there are available actions. Some behavioral biases can be captured by the initial conditions on these hidden value states.

- a learning rate. This parameter controls the impact of reward prediction error on the value update. Note that one may want to ask whether the learning rate depends upon experimental factors (pathological condition, gain/loss domains, etc...)

- a behavioral temperature and a bias. These parameters control the exploitation vs exploration ratio of the agent, and capture choice randomness.

Q-learning models simply assume that subjects update the value of possible actions. In its simplest form, the Q-learning algorithm expresses the change in value $Q(t+1)-Q(t)$ from trial t to trial $t+1$ as being linearly proportional to the prediction error. This yields the following learning rule:

$$Q_{t+1} = Q_t + \alpha * (r_{t+1} - Q_t)$$

Where r_t is the reward delivered to the subject at trial t , and α is the (unknown) "learning rate" of the subject.

The Q-learning evolution function thus takes the agent's previous action and the feedback received for the previous action as data inputs. We complement with a softmax decision rule in the Q-learning observation function. As a reminder, this is an equation that expresses the probability $P_t(a_i)$ of the subject to choose action a_i at trial t :

$$P_t(a_i) = \frac{\exp(\beta * Q_t(a_i))}{\sum_j \beta * Q_t(a_j)}$$

Where β is the (unknown) temperature.

Fitting the above Q-learning model to behavioral data means finding estimates of the learning rate α , the inverse temperature β , and the initial values Q_0 that best explains the observed choices.

Influence Learning model

This mentalizing model was described by Hampton et al¹⁷, who also found neural correlates with this model in humans¹⁷. The influence model relies on a Taylor expanded reinforcement learning. This model describes computations underlying the capacity to mentalize (in the context of a strategic game). Let $p_n = P_n(o = 1)$ be the agent's prediction of the other's next move, i.e., the probability that the other will pick the first alternative option. The "influence learning" rule can be written as follows:

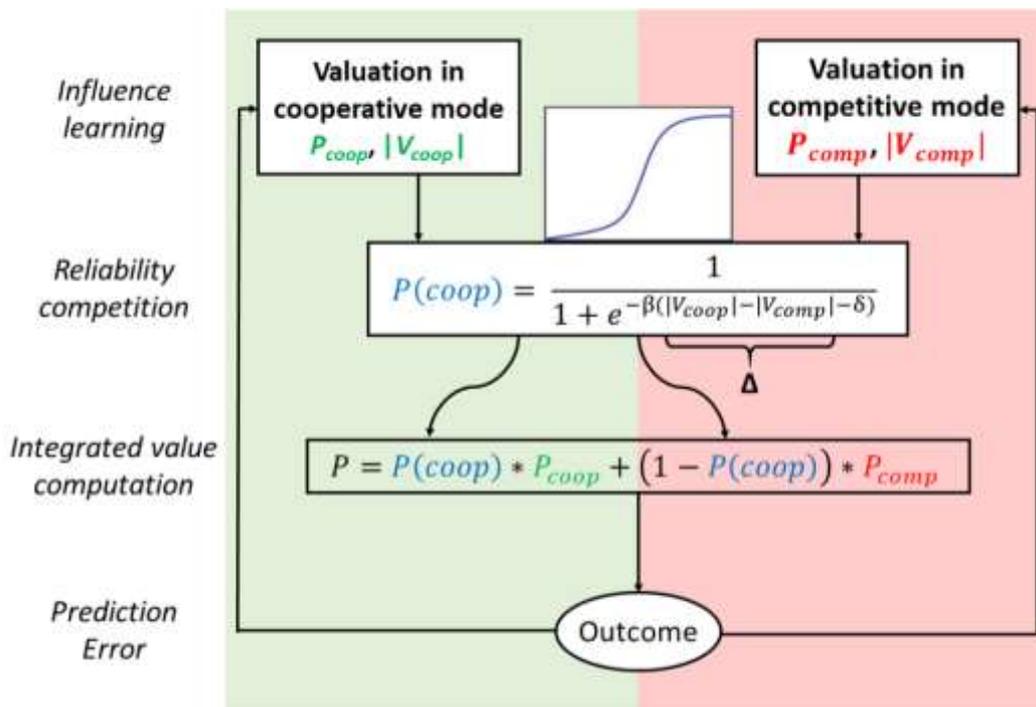
$$p_{n+1} = p_n + \eta * (o - p_n) - \lambda * k_1 * p_n * (1 - p_n) * \left(a - \frac{\beta * x_n - k_2}{k_1}\right) \quad (1)$$

where o is the other's last move, a is the agent's last move, η is the weight of the agent's prediction error, λ is the weight of the other's prediction error, k_1 & k_2 are derived from the game's payoff table, x_n is the tracked log-odds of p_n , β is the other's temperature, i.e. the tendency of the other to explore vs exploit information (the closer it is to 0, the more it explores by choosing randomly without using previous information, and the closer it is 1, the more it exploits previous information and chooses deterministically).

Note that for all subjects, the tested models fitted better than chance.

Mixed-Intentions Influence Learning (MIIL) model

The 'mixed intentions' version of the influence model computes one decision value according to a competitive expert and another according to a cooperative expert and then arbitrates between the two, based on the difference in their respective reliability (see SI, Fig. 1). We defined reliability as the difference in unsigned value functions for two choices given by particular learning algorithms.



Supp. Fig.1. Scheme of the Mixed-Intention Influence model (adapted from Philippe et al¹²). Two influence models (one cooperative and the other competitive, resp. P_{coop} in green and P_{comp} in red) compute a value for choosing one specific target (P in black). A controller uses the difference between the absolute value of the value of each expert (called reliability) to compute a probability that the other is cooperating (blue). Then, the model weights the value of each expert according to the probability of being in cooperative or in competitive modes to produce a final decision value. Then it compares its predictions to the actual reward and computes again a new value for each expert.

GLMs & GLMEs.

We at first modeled our GLM based on the following formulas:

$\text{Reward}_n \sim 1 + \text{Reward}_{n-1} + \text{Reward}_{n-2} + \text{Reward}_{n-3} + \text{Stay}_{n-1} + \text{Stay}_{n-2} + \text{Stay}_{n-3} + \text{AA_Stay}_{n-1} + \text{AA_Stay}_{n-2} + \text{AA_Stay}_{n-3} + \text{Coop_sig}_{n-1} + \text{Coop_sig}_{n-2} + \text{Trial} + \text{AA_mode} + \text{Hierar_cond} + \text{Sex} + \text{Age} + \text{DT} + (\text{Age}:\text{Coop_sig}_{n-1}) + (\text{Age}:\text{Coop_sig}_{n-2}) + (\text{Age}:\text{AA_mode})$

$\text{Stay}_n \sim 1 + \text{Reward}_{n-1} + \text{Reward}_{n-2} + \text{Reward}_{n-3} + \text{Stay}_{n-1} + \text{Stay}_{n-2} + \text{Stay}_{n-3} + \text{AA_Stay}_{n-1} + \text{AA_Stay}_{n-2} + \text{AA_Stay}_{n-3} + \text{Coop_sig}_{n-1} + \text{Coop_sig}_{n-2} + \text{Trial} + \text{AA_mode} + \text{Hierar_cond} + \text{Sex} + \text{Age} + \text{DT} + (\text{Age}:\text{Coop_sig}_{n-1}) + (\text{Age}:\text{Coop_sig}_{n-2}) + (\text{Age}:\text{AA_mode})$

The R summary outputs were:

```
glm(formula = reward_n ~ 1 + reward_n_1 + reward_n_2 + reward_n_3 +
     stay_n_1 + stay_n_2 + stay_n_3 + AI_stay_n_1 + AI_stay_n_2 +
     AI_stay_n_3 + coop_sig_n + coop_sig_n_1 + coop_sig_n:Age +
     coop_sig_n_1:Age + Trial + Age + mode + mode:Age + Sex +
     Condition + DT_interval, family = binomial, data = all_data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6512	-1.1944	0.8336	1.1220	1.4958

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.882495	0.166709	-5.294	1.20e-07 ***
reward_n_1	0.135809	0.030741	4.418	9.97e-06 ***
reward_n_2	0.224299	0.031396	7.144	9.05e-13 ***
reward_n_3	0.095392	0.031254	3.052	0.002272 **
stay_n_1	0.319626	0.030827	10.368	< 2e-16 ***
stay_n_2	0.006183	0.040066	0.154	0.877365
stay_n_3	0.084637	0.039147	2.162	0.030617 *
AI_stay_n_1	0.107068	0.040205	2.663	0.007744 **
AI_stay_n_2	0.077600	0.039470	1.966	0.049290 *
AI_stay_n_3	0.103181	0.028713	3.594	0.000326 ***
coop_sig_n	-0.241575	0.241846	-0.999	0.317853
coop_sig_n_1	0.302307	0.186278	1.623	0.104615
Trial	0.045286	0.015080	3.003	0.002673 **
Age	0.048033	0.024527	1.958	0.050183 .
mode	-0.040009	0.134166	-0.298	0.765548
SexMale	-0.026873	0.028118	-0.956	0.339217
Condition	-0.027190	0.021363	-1.273	0.203097
DT_interval	0.002459	0.014542	0.169	0.865699
coop_sig_n:Age	0.047194	0.036310	1.300	0.193687
coop_sig_n_1:Age	-0.029094	0.028008	-1.039	0.298920
Age:mode	0.011641	0.020080	0.580	0.562078

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 29242 on 21229 degrees of freedom
Residual deviance: 28563 on 21209 degrees of freedom
(772 observations deleted due to missingness)
AIC: 28605

Number of Fisher Scoring iterations: 4

```
glm(formula = stay_n ~ 1 + reward_n_1 + reward_n_2 + reward_n_3 +  
  stay_n_1 + stay_n_2 + stay_n_3 + AI_stay_n_1 + AI_stay_n_2 +  
  AI_stay_n_3 + coop_sig_n + coop_sig_n_1 + coop_sig_n:Age +  
  coop_sig_n_1:Age + Trial + Age + mode + mode:Age + Sex +  
  Condition + DT_interval, family = binomial, data = all_data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7896	-1.1789	0.7449	1.0952	1.5737

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.921750	0.167804	-5.493	3.95e-08 ***
reward_n_1	0.427364	0.030985	13.793	< 2e-16 ***
reward_n_2	-0.005308	0.031660	-0.168	0.86684
reward_n_3	0.080017	0.031511	2.539	0.01111 *
stay_n_1	0.125316	0.031079	4.032	5.53e-05 ***
stay_n_2	0.189470	0.040200	4.713	2.44e-06 ***
stay_n_3	0.124653	0.039408	3.163	0.00156 **
AI_stay_n_1	0.196072	0.040380	4.856	1.20e-06 ***
AI_stay_n_2	0.171727	0.039696	4.326	1.52e-05 ***
AI_stay_n_3	0.252993	0.028970	8.733	< 2e-16 ***
coop_sig_n	0.004318	0.243123	0.018	0.98583
coop_sig_n_1	-0.095439	0.186746	-0.511	0.60931
Trial	0.072417	0.015296	4.734	2.20e-06 ***
Age	0.006792	0.024671	0.275	0.78309
mode	-0.130648	0.134809	-0.969	0.33248
SexMale	-0.044579	0.028475	-1.566	0.11746
Condition	-0.034473	0.021639	-1.593	0.11114
DT_interval	-0.014021	0.014684	-0.955	0.33967
coop_sig_n:Age	0.031188	0.036493	0.855	0.39276
coop_sig_n_1:Age	0.038038	0.028066	1.355	0.17532
Age:mode	0.020517	0.020172	1.017	0.30911

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 29199 on 21229 degrees of freedom
Residual deviance: 28026 on 21209 degrees of freedom
(772 observations deleted due to missingness)
AIC: 28068

Number of Fisher Scoring iterations: 4

After applying a dropterm() process based on the AICs of the models successively, the final models were:

```
Reward_n ~ 1 + Reward_{n-1} + Reward_{n-2} + Reward_{n-3} + Stay_{n-1} + Stay_{n-3} + AA_Stay_{n-1} +  
AA_Stay_{n-2} + AA_Stay_{n-3} + Coop_sig_{n-1} + Coop_sig_{n-2} + Trial + Age + (Age:Coop_sig_{n-1})  
Stay_n ~ 1 + Reward_{n-1} + Reward_{n-3} + Stay_{n-1} + Stay_{n-2} + Stay_{n-3} + AA_Stay_{n-1} + AA_Stay_{n-2} +  
AA_Stay_{n-3} + coop_sig_{n-1} + coop_sig_{n-2} + Trial + Age + AA_mode + (AA_mode:Age) +  
(Age:coop_sig_{n-2}) + Sex + Condition
```

We used the glm() function of the lme4 package on R with a binomial family.

The R summary outputs were:

```
glm(formula = reward_n ~ 1 + reward_n_1 + reward_n_2 + reward_n_3 +  
  stay_n_1 + stay_n_3 + AI_stay_n_1 + AI_stay_n_2 + AI_stay_n_3 +  
  coop_sig_n + coop_sig_n_1 + coop_sig_n:Age + Trial + Age,  
  family = binomial, data = all_data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.6400	-1.1939	0.8343	1.1230	1.4677

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.79419	0.11351	-6.997	2.62e-12	***
reward_n_1	0.13777	0.03067	4.492	7.04e-06	***
reward_n_2	0.22214	0.03086	7.197	6.14e-13	***
reward_n_3	0.09386	0.03117	3.011	0.002603	**
stay_n_1	0.31885	0.03078	10.358	< 2e-16	***
stay_n_3	0.08633	0.03907	2.210	0.027134	*
AI_stay_n_1	0.11296	0.03089	3.657	0.000255	***
AI_stay_n_2	0.07616	0.03940	1.933	0.053244	.
AI_stay_n_3	0.10179	0.02862	3.557	0.000375	***
coop_sig_n	-0.19871	0.13376	-1.486	0.137403	
coop_sig_n_1	0.13516	0.03963	3.410	0.000649	***
Trial	0.04219	0.01474	2.863	0.004191	**
Age	0.02797	0.01644	1.701	0.088926	.
coop_sig_n:Age	0.04749	0.01978	2.401	0.016364	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 29242 on 21229 degrees of freedom
Residual deviance: 28568 on 21216 degrees of freedom
(772 observations deleted due to missingness)
AIC: 28596

Number of Fisher Scoring iterations: 4

```
glm(formula = stay_n ~ 1 + reward_n_1 + reward_n_3 + stay_n_1 +  
  stay_n_2 + stay_n_3 + AI_stay_n_1 + AI_stay_n_2 + AI_stay_n_3 +  
  coop_sig_n + coop_sig_n_1 + coop_sig_n_1:Age + Trial + Age +  
  mode + mode:Age + Sex + Condition, family = binomial, data = all_data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7870	-1.1789	0.7452	1.0949	1.5906

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.01799	0.12891	-7.897	2.86e-15	***
reward_n_1	0.42592	0.03092	13.774	< 2e-16	***
reward_n_3	0.08087	0.03148	2.569	0.010205	*
stay_n_1	0.12632	0.03096	4.080	4.50e-05	***
stay_n_2	0.18913	0.03966	4.769	1.86e-06	***

```

stay_n_3      0.12491  0.03939  3.171 0.001520 **
AI_stay_n_1   0.19588  0.03974  4.929 8.28e-07 ***
AI_stay_n_2   0.17140  0.03960  4.328 1.50e-05 ***
AI_stay_n_3   0.25365  0.02896  8.758 < 2e-16 ***
coop_sig_n    0.20653  0.05734  3.602 0.000316 ***
coop_sig_n_1  -0.11472  0.18523  -0.619 0.535695
Trial         0.07281  0.01528  4.764 1.89e-06 ***
Age           0.02112  0.01835  1.151 0.249580
mode         -0.20953  0.09820  -2.134 0.032871 *
SexMale      -0.04353  0.02845  -1.530 0.126007
Condition    -0.03278  0.02158  -1.519 0.128665
coop_sig_n_1:Age 0.04109  0.02781  1.477 0.139571
Age:mode     0.03259  0.01429  2.280 0.022580 *

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 29199 on 21229 degrees of freedom
Residual deviance: 28028 on 21212 degrees of freedom
(772 observations deleted due to missingness)
AIC: 28064

Number of Fisher Scoring iterations: 4

The successive factors that were dropped by our GLM dropterm() procedure in R were:

Reward _n (GLM)	Df	Deviance	AIC
stay_n_2	1	28562.75	28602.75
DT_interval	1	28562.77	28600.77
Age:mode	1	28563.11	28599.11
coop_sig_n_2:Age	1	28563.88	28597.88
Sex	1	28564.81	28596.81
Mode	1	28566.08	28596.08
Condition	1	28567.89	28595.89
<none>	NA	28567.89	28595.89

Stay _n (GLM)	Df	Deviance	AIC
reward_n_2	1	28026.14	28066.14
coop_sig_n_1:Age	1	28026.87	28064.87
DT_interval	1	28027.77	28063.77
<none>	NA	28027.77	28063.77

We then checked if our results were still robust if we included random effects of time. We used the glmer() function of the lme4 package on R with a binomial family and the bobyqa optimizer. We proceeded to further simplifications through another dropterm() process which led to the following formula:

Reward_n ~ 1 + Reward_{n-1} + Reward_{n-2} + Reward_{n-3} + Stay_{n-1} + Stay_{n-3} + AA_Stay_{n-1} + AA_Stay_{n-2} + AA_Stay_{n-3} + Coop_sig_{n-1} + Coop_sig_{n-2} + Trial + Age + (Age:Coop_sig_{n-1}) + (1 + Trial | ID)

Stay_n ~ 1 + Reward_{n-1} + Reward_{n-3} + Stay_{n-2} + AA_Stay_{n-1} + AA_Stay_{n-2} + AA_Stay_{n-3} + Coop_sig_{n-1} + Coop_sig_{n-2} + Trial + Age + AA_mode + (AA_mode:Age) + (Age:Coop_sig_{n-2}) + (1 + Trial | ID)

Their summary outputs in R were:

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
['glmerMod']

Family: binomial (logit)

Formula: reward_n ~ 1 + reward_n_1 + reward_n_2 + reward_n_3 + stay_n_1 + stay_n_3 + AI_stay_n_1 + AI_stay_n_2 + AI_stay_n_3 + coop_sig_n_1 + coop_sig_n_2 + coop_sig_n_1:Age + Trial + Age + (1 + Trial | ID)

Data: all_data
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))

AIC	BIC	logLik	deviance	df.resid
28570.6	28706.0	-14268.3	28536.6	21213

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.9889	-1.0158	0.6215	0.9348	1.3948

Random effects:

Groups Name	Variance	Std.Dev.	Corr
ID (Intercept)	0.029156	0.17075	
Trial	0.001403	0.03746	0.54

Number of obs: 21230, groups: ID, 193

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.71923	0.12685	-5.670	1.43e-08	***
reward_n_1	0.11649	0.03120	3.734	0.000188	***
reward_n_2	0.20311	0.03132	6.485	8.86e-11	***
reward_n_3	0.07641	0.03158	2.420	0.015529	*
stay_n_1	0.30152	0.03122	9.659	< 2e-16	***
stay_n_3	0.07109	0.03944	1.803	0.071428	.
AI_stay_n_1	0.09678	0.03129	3.094	0.001978	**
AI_stay_n_2	0.06985	0.03958	1.765	0.077630	.
AI_stay_n_3	0.07906	0.02918	2.710	0.006737	**
coop_sig_n_1	-0.21691	0.13435	-1.614	0.106423	.
coop_sig_n_2	0.13944	0.03975	3.508	0.000452	***
Trial	0.04419	0.01506	2.935	0.003341	**
Age	0.02724	0.01839	1.481	0.138522	.
coop_sig_n:Age	0.05124	0.01988	2.577	0.009963	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation matrix not shown by default, as $p = 14 > 12$.
Use `print(x, correlation=TRUE)` or
`vcov(x)` if you need it

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)

['glmerMod']

Family: binomial (logit)

Formula: `stay_n ~ 1 + reward_n_1 + reward_n_3 + stay_n_2 + AI_stay_n_1 + AI_stay_n_2 + AI_stay_n_3 + coop_sig_n_1 + coop_sig_n_2 + coop_sig_n_2:Age + Trial + Age + mode + mode:Age + (1 + Trial | ID)`

Data: all_data

Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+05))

AIC	BIC	logLik	deviance	df.resid
27812.0	27947.4	-13889.0	27778.0	21213

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.5128	-0.9796	0.4887	0.8999	1.8841

Random effects:

Groups Name Variance Std.Dev. Corr
 ID (Intercept) 0.12670 0.3559
 Trial 0.03475 0.1864 0.12
 Number of obs: 21230, groups: ID, 193

Fixed effects:

 Estimate Std. Error z value Pr(>|z|)
 (Intercept) -0.78275 0.16996 -4.605 4.12e-06 ***
 reward_n_1 0.42514 0.02938 14.468 < 2e-16 ***
 reward_n_3 0.08482 0.03163 2.681 0.007329 **
 stay_n_2 0.09117 0.04080 2.235 0.025437 *
 AI_stay_n_1 0.16911 0.04046 4.180 2.92e-05 ***
 AI_stay_n_2 0.16380 0.03174 5.161 2.45e-07 ***
 AI_stay_n_3 0.16712 0.02984 5.601 2.13e-08 ***
 coop_sig_n_1 0.22127 0.05848 3.784 0.000155 ***
 coop_sig_n_2 -0.12465 0.18971 -0.657 0.511133
 Trial 0.08319 0.02059 4.041 5.32e-05 ***
 Age 0.01886 0.02480 0.760 0.447049
 mode -0.24768 0.10027 -2.470 0.013507 *
 coop_sig_n_2:Age 0.04671 0.02853 1.637 0.101541
 Age:mode 0.03843 0.01461 2.630 0.008539 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The successive factors that were dropped by our GLMER dropterm() procedure in R were:

Reward _n (GLMER)	Df	Deviance	AIC
<none>	NA	28536.6	28570.6

Stay _n (GLMER)	Df	Deviance	AIC
stay_n_1	1	27776.5	27816.5
stay_n_3	1	27776.7	27814.7
Sex	1	27777.1	27813.1
Condition	1	27778.0	27812.0
<none>	NA	27778.0	27812.0

We then proceeded to an addterm() process with the regressors that were previously removed from the GLMs:

add_term_score_GLMER = addterm(GLMER_score, ~ . + Stay_{n-2} + (Age:Coop_sig_{n-2}) + mode + Sex + Condition + DT, test="Chisq")

add_term_stay_GLMER = addterm(GLMER_stay, ~ . + Reward_{n-2} + (Age:Coop_sig_{n-1}) + Sex + Condition + DT, test="Chisq")

This led to the following formula and results:

Reward_n ~ 1 + Reward_{n-1} + Reward_{n-2} + Reward_{n-3} + Stay_{n-1} + Stay_{n-3} + AA_Stay_{n-1} + AA_Stay_{n-2} + AA_Stay_{n-3} + Coop_sig_{n-1} + Coop_sig_{n-2} + Trial + Age + (Age:Coop_sig_{n-1}) + (1 + Trial |ID)

Stay_n ~ 1 + Reward_{n-1} + Reward_{n-3} + Stay_{n-2} + AA_Stay_{n-1} + AA_Stay_{n-2} + AA_Stay_{n-3} + Coop_sig_{n-1} + Coop_sig_{n-2} + Trial + Age + AA_mode + DT + (AA_mode:Age) + (Age:Coop_sig_{n-2}) + (1 + Trial |ID)

Models	GLM		GLME	
Variables	Reward _n : Estimate (Std. Error)	Stay _n : Estimate (Std. Error)	Reward _n : Estimate (Std. Error)	Stay _n : Estimate (Std. Error)
Intercept	-0.79419 (0.11351) ***	-1.01799 (0.12891) ***	-0.71931 (0.12678) ***	-0.78958 (0.17080) ***
Reward _{n-1}	0.13777 (0.03067) ***	0.42592 (0.03092) ***	0.11649 (0.03120) ***	0.42246 (0.02944) ***
Reward _{n-2}	0.22214 (0.03086) ***	/	0.20311 (0.03132) ***	/
Reward _{n-3}	0.09386 (0.03117) **	0.08087 (0.03148) *	0.07641 (0.03158) *	0.08593 (0.03164) **
Stay _{n-1}	0.31885 (0.03078) ***	0.12632 (0.03096) ***	0.30152 (0.03122) ***	/
Stay _{n-2}	/	0.18913 (0.03966) ***	/	0.09100 (0.04080) *
Stay _{n-3}	0.08633 (0.03907) *	0.12491 (0.03939) **	0.07109 (0.03943) .	/
AA_Stay _{n-1}	0.11296 (0.03089) ***	0.19588 (0.03974) ***	0.09678 (0.03128) **	0.17040 (0.04047) ***
AA_Stay _{n-2}	0.07616 (0.03940) .	0.17140 (0.03960) ***	0.06986 (0.03958) .	0.16351 (0.03174) ***
AA_Stay _{n-3}	0.10179 (0.02862) ***	0.25365 (0.02896) ***	0.07905 (0.02918) **	0.16725 (0.02984) ***
Coop_sig _{n-1}	-0.19871 (0.13376)	0.20653 (0.05734) ***	-0.21677 (0.13427)	0.22357 (0.05851) ***
Coop_sig _{n-2}	0.13516 (0.03963) ***	-0.11472 (0.18523)	0.13944 (0.03975) ***	-0.12383 (0.18964)
Trial	0.04219 (0.01474) **	0.07281 (0.01528) ***	0.04419 (0.01506) **	0.08420 (0.02062) ***
Age	0.02797 (0.01644) .	0.02112 (0.01835)	0.02725 (0.01838)	0.01987 (0.02493)
AA_mode	/	-0.20953 (0.09820) *	/	-0.24844 (0.10025) *
Sex (Male)	/	-0.04353 (0.02845)	/	/
Hierar_cond	/	-0.03278 (0.02158)	/	/
DT	/	/	/	0.02368 (0.01623)
Coop_sig _{n-1} :Age	0.04749 (0.01978) *	/	0.05122 (0.01987) **	/
Coop_sig _{n-2} :Age	/	0.04109 (0.02781)	/	0.04668 (0.02852)
AA_mode:Age	/	0.03259 (0.01429) *	/	0.03831 (0.01461) **

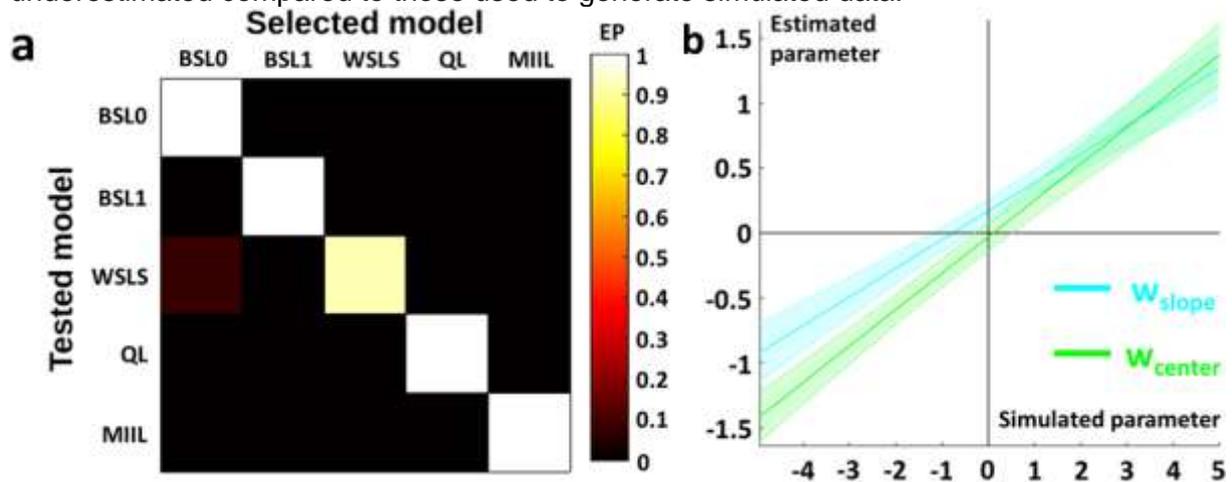
Supp. Table 1: Comparison of GLMs and GLMEs for behavioral factors. Hierar_cond, AA_mode, Coop_sig, Reward, AA_Stay and Stay are coded effects (Hierar_cond: different = 1, equal = 0 ; AA_mode: competitive = -1, random = 0, cooperative = 1; Coop_sig, Reward, AA_Stay & Stay: True = 1, False = 0). Trial & DT were normalized. Signif. codes: '***': p < 0.001; '**': p < 0.01; '*': p < 0.05.

Control analyses and model checks

We ran several control analyses. We first conducted a Confusion Analysis (CA) as a model check to ensure that the compared models that we tested, given our data, were not confused between each other and that our model selection was indeed reliable⁴¹. We therefore generated data multiple times with Monte-Carlo simulations (approx. 2800 in total) for each of our tested models with different sets of parameters, and conducted the same analysis as

presented previously on the simulated data. This meant fitting the data with all of the tested models, then performing a BMS on the models to select the one that fitted best. If this analysis is reliable and the models are distinguishable, then the model that we simulated at the beginning should also be that selected by our analysis. The confusion matrix (Supp.Fig.2.a, see also the confusion table in Supp. Mat.) showed that the final selected model was indeed that which was initially simulated the majority of the time. Even if data generated by all models, the MIIL model included, can, to a certain extent, be mistakenly attributed to the BSL0 model, the reciprocal is not true, i.e. data generated by a BSL0 is not mistaken for any other model, the MIIL model included. This means that we perhaps actually underestimate the amount of AToM in younger children.

Next, we checked if the learning parameters of the MIIL model could actually be identified from observed data. To do so, we generated simulated data with multiple Monte-Carlo simulations (approx. 9800 in total) from the influence model using different sets of learning parameters sampled from a uniform distribution. The sampling was done (uniformly) between the most extreme empirical values found in our data analysis for each of our seven learning parameters. We then fitted the generated data with the influence model with the same method as previously. We then regressed the estimated parameters on the simulated parameters. We found for each learning parameter a positive correlation between estimated and simulated parameters (Fig.3.b). The regression indicates that the estimated parameters were actually underestimated compared to those used to generate simulated data.

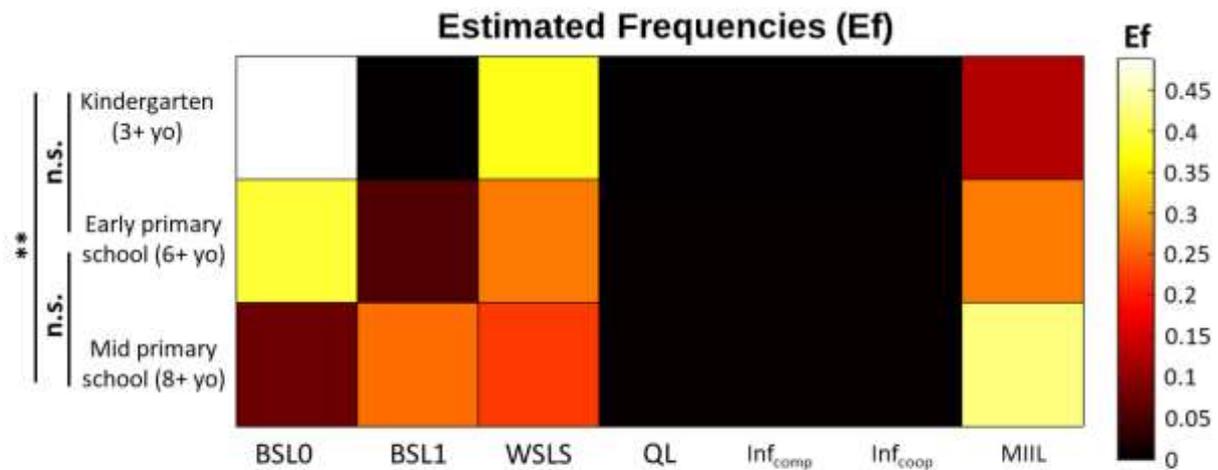


Supp. Fig. 2. Model checks. (a) Confusion matrix. x-axis: model selected as the best fitting one after model fitting and Bayesian Model Selection (BMS); y-axis: initial simulated model that generated the data; Color-scale: model attribution ratio. **(b) Simulation-recovery analysis.** X-axis: value of the simulated parameters of the influence model that generated the simulated data; Y-axis: value of the estimated parameter after model fitting and parameter estimation on the simulated data. We simulated then recovered the two parameters constituting the arbitration weight that estimates the cooperation tendency of the subject in the MIIL model.

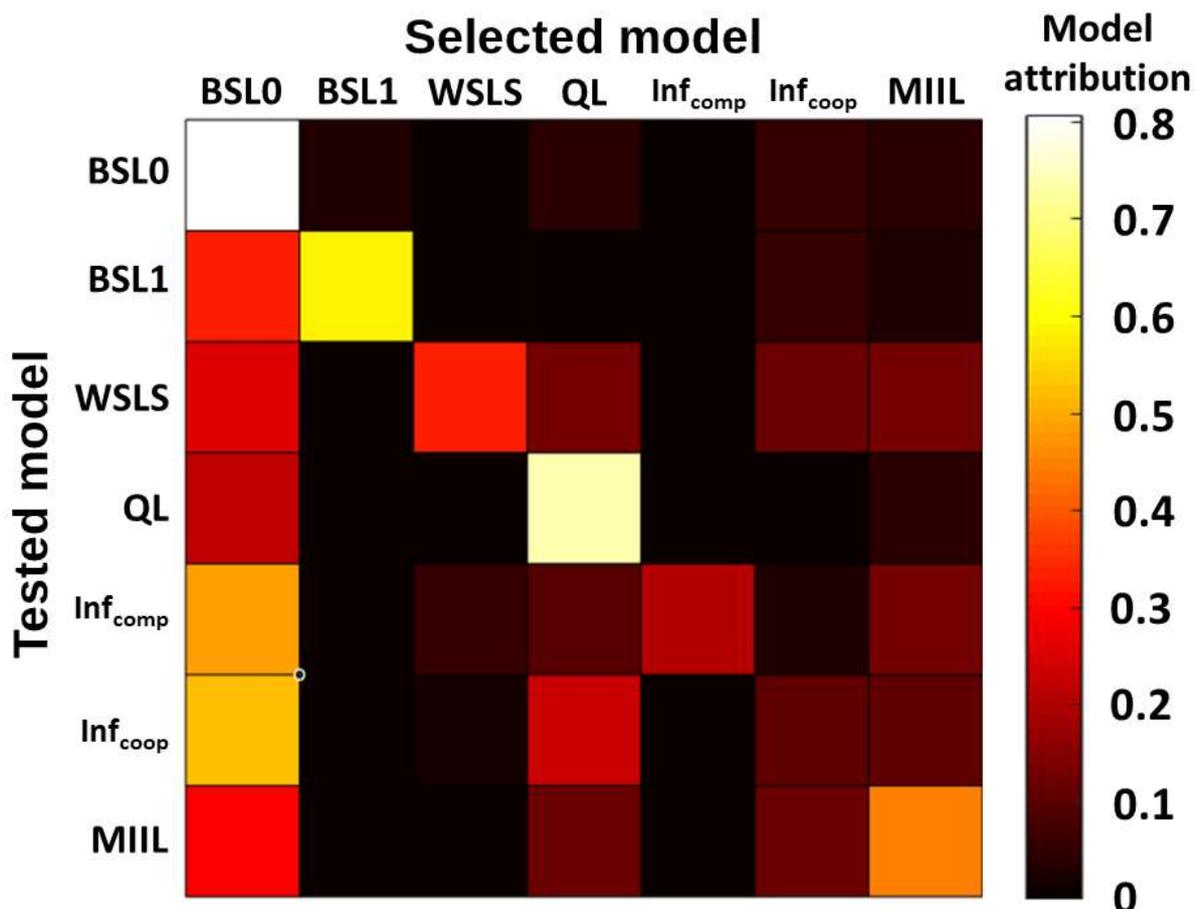
Supplementary analyses

We initially fitted and compared other models than those presented in the main results: 2 influence models, one in a purely competitive setup and one in a purely cooperative setup. But they eventually did not pass the confusion analysis, i.e. they were confused with other models when simulated then tested through our analysis. They indeed get confused for the BSL0 model. This implies that, for our main results, the overrepresentation of the BSL0 model in the younger groups might actually hide the presence of cooperative and competitive Influence models among them. However, this does not change the main result that a transition

in model representation occurs around age 8 in MPS when the MIIL model becomes preponderant. Still, the results of the comparison is given here, for the readers' information.



P=0,0067
(uncorrected)



Supp. Fig. 3. Supplementary analysis. (top) Bayesian model comparison. X-axis: model names; Y-axis: age group; Color-scale: Estimated frequencies (Ef), i.e., the frequency at which one model fitted one subject best in the group population compared to the other tested models. In the Kindergarten (KG) and Early Primary School (EPS) groups, the model that best fitted most of our group of participants is the BSL0. In Mid-Primary School, the MIIL model best fitted

most of the participants. **(down) Confusion matrix.** X-axis: model selected as the best fitting one after model fitting and Bayesian Model Selection (BMS); Y-axis: initial simulated model that generated the data; Color-scale: model attribution ratio.

Bibliography

1. Premack, D. & Woodruff, G. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* **1**, 515–526 (1978).
2. Rakoczy, H. Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology* **2022 1:4 1**, 223–235 (2022).
3. Ho, M. K., Saxe, R. & Cushman, F. Planning with Theory of Mind. *Trends Cogn Sci* **26**, 959–971 (2022).
4. Miller, S. A. Children’s Understanding of Second-Order Mental States. *Psychol Bull* **135**, 749–773 (2009).
5. Osterhaus, C. & Koerber, S. The Development of Advanced Theory of Mind in Middle Childhood: A Longitudinal Study From Age 5 to 10 Years. *Child Dev* **92**, 1872–1888 (2021).
6. Osterhaus, C., Koerber, S. & Sodian, B. Scientific thinking in elementary school: Children’s social cognition and their epistemological understanding promote experimentation skills. *Dev Psychol* **53**, 450–462 (2017).
7. Osterhaus, C. & Koerber, S. Social cognition during and after kindergarten: The relations between first-order and advanced theories of mind. *European Journal of Developmental Psychology* **18**, 573–592 (2021).
8. Osterhaus, C., Koerber, S. & Sodian, B. Scaling of Advanced Theory-of-Mind Tasks. *Child Dev* **87**, 1971–1991 (2016).
9. Koerber, S. & Osterhaus, C. Some but not all aspects of (advanced) theory of mind predict loneliness. *British Journal of Developmental Psychology* **38**, 144–148 (2020).
10. Hayward, E. O. & Homer, B. D. Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology* **35**, 454–462 (2017).
11. Hughes, C. *et al.* Good Test–Retest Reliability for Standard and Advanced False-Belief Tasks across a Wide Range of Abilities. *J. Child Psychol. Psychiat.* **41**, 483–490 (2000).
12. Philippe, R. *et al.* Neurocomputational mechanisms engaged in detecting cooperative and competitive intentions of others. *In Revision* 0–42 (2022) doi:10.21203/rs.3.rs-1160167/v1.
13. Klindt, D., Devaine, M. & Daunizeau, J. Does the way we read others’ mind change over the lifespan? Insights from a massive web poll of cognitive skills from childhood to late adulthood. *Cortex* **86**, 205–215 (2017).
14. Devaine, M., Hollard, G. & Daunizeau, J. Theory of mind: Did evolution fool us? *PLoS One* **9**, (2014).
15. Devaine, M., Hollard, G. & Daunizeau, J. The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? *PLoS Comput Biol* **10**, (2014).

16. Rusch, T., Steixner-Kumar, S., Doshi, P., Spezio, M. & Gläscher, J. Theory of mind and decision science: Towards a typology of tasks and computational models. *Neuropsychologia* **146**, 1–44 (2020).
17. Hampton, A. N., Bossaerts, P. & O’Doherty, J. P. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci U S A* **105**, 6741–6746 (2008).
18. Hill, C. A. *et al.* A causal account of the brain network computations underlying strategic social behavior. *Nat Neurosci* **20**, 1142–1149 (2017).
19. Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner’s Dilemma game. *Nature* **364**, 56–58 (1993).
20. O’Doherty, J. P., Cockburn, J. & Pauli, W. M. Learning, Reward, and Decision Making. *Annual Review of Psychology* vol. 68 73–100 Preprint at <https://doi.org/10.1146/annurev-psych-010416-044216> (2017).
21. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* **5**, 9 (2011).
22. Mathys, C. D. *et al.* Uncertainty in perception and the Hierarchical Gaussian filter. *Front Hum Neurosci* **8**, 1–24 (2014).
23. Trepka, E. *et al.* Entropy-based metrics for predicting choice behavior based on local response to reward. *Nat Commun* **12**, 1–16 (2021).
24. Woo, J. H. *et al.* Mechanisms of adjustments to different types of uncertainty in the reward environment across mice and monkeys. *Cogn Affect Behav Neurosci* **23**, 600–619 (2023).
25. Rigoux, L., Stephan, K. E., Friston, K. J. & Daunizeau, J. Bayesian model selection for group studies - revisited. *Neuroimage* **84**, 971–985 (2014).
26. Daunizeau, J. & Rigoux, L. Bayesian sequence learning. <https://mbb-team.github.io/VBA-toolbox/wiki/BSL/> <https://mbb-team.github.io/VBA-toolbox/wiki/BSL/> (2014).
27. Daunizeau, J. & Rigoux, L. The variational Bayes approach. <https://mbb-team.github.io/VBA-toolbox/wiki/The-variational-Bayesian-approach/> (2014).
28. Daunizeau, J., Adam, V. & Rigoux, L. VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Comput Biol* **10**, 1003441 (2014).
29. Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J. & Friston, K. J. Bayesian model selection for group studies. *Neuroimage* **46**, 1004–1017 (2009).
30. Devine, R. T. & Hughes, C. Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *J Exp Child Psychol* **149**, 23–40 (2016).
31. Charpentier, C. J., Iigaya, K. & O’Doherty, J. P. A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. *Neuron* **106**, 687–699.e7 (2020).

32. Kyriakopoulou, N. & Vosniadou, S. Theory of Mind, Personal Epistemology, and Science Learning: Exploring Common Conceptual Components. *Front Psychol* **11**, 1140 (2020).
33. Astington, J. W., Pelletier, J. & Homer, B. Theory of mind and epistemological development: the relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas Psychol* **20**, 131–144 (2002).
34. Sher, I., Koenig, M. & Rustichini, A. Children's strategic theory of mind. *Proc Natl Acad Sci U S A* **111**, 13307–13312 (2014).
35. Fink, E., Begeer, S., Peterson, C. C., Slaughter, V. & de Rosnay, M. Friendlessness and theory of mind: A prospective longitudinal study. *British Journal of Developmental Psychology* **33**, 1–17 (2015).
36. Bosacki, S. L. Children's theory of mind, self-perceptions, and peer relations: A longitudinal study. *Infant Child Dev* **24**, 175–188 (2015).
37. Bosacki, S., Moreira, F. P., Sitnik, V., Andrews, K. & Talwar, V. Theory of Mind, Self-Knowledge, and Perceptions of Loneliness in Emerging Adolescents. *Journal of Genetic Psychology* **181**, 14–31 (2020).
38. Weinstock, M., Israel, V., Cohen, H. F., Tabak, I. & Harari, Y. Young Schoolchildren's Epistemic Development: A Longitudinal Qualitative Study. *Front Psychol* **11**, 1475 (2020).
39. Lecce, S., Bianco, F., Devine, R. T. & Hughes, C. Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *J Exp Child Psychol* **163**, 69–86 (2017).
40. Ebert, S. Theory of mind, language, and reading: Developmental relations from early childhood to early adolescence. *J Exp Child Psychol* **191**, 104739 (2020).
41. Palminteri, S., Wyart, V. & Koechlin, E. The Importance of Falsification in Computational Cognitive Modeling. *Trends in Cognitive Sciences* vol. 21 425–433 Preprint at <https://doi.org/10.1016/j.tics.2017.03.011> (2017).

