Supplementary information and analyses for:

Moving developmental research online: comparing in-lab and web-based studies of model-based reinforcement learning

Participant recruitment process

Data collection took place from June 1 - July 10, 2020. Each week, we determined target recruitment numbers based on ensuring an even spread of participants across age and gender bins (Figure S1). We then emailed eligible participants a description of the study with a personalized, single-use link to a Qualtrics consent form. Each participant's link embedded their study ID number, which would ultimately get carried through to the link that launched the first experimental task and saved in their data file. Qualtrics links expired one week from the date we sent them, enabling us to continuously track how many completed, active, and expired links had been sent to participants from each age/gender bin, and adjust our recruitment goals for the next week accordingly. Participants were sent reminder emails three days after their initial recruitment email if they had not yet completed the study.

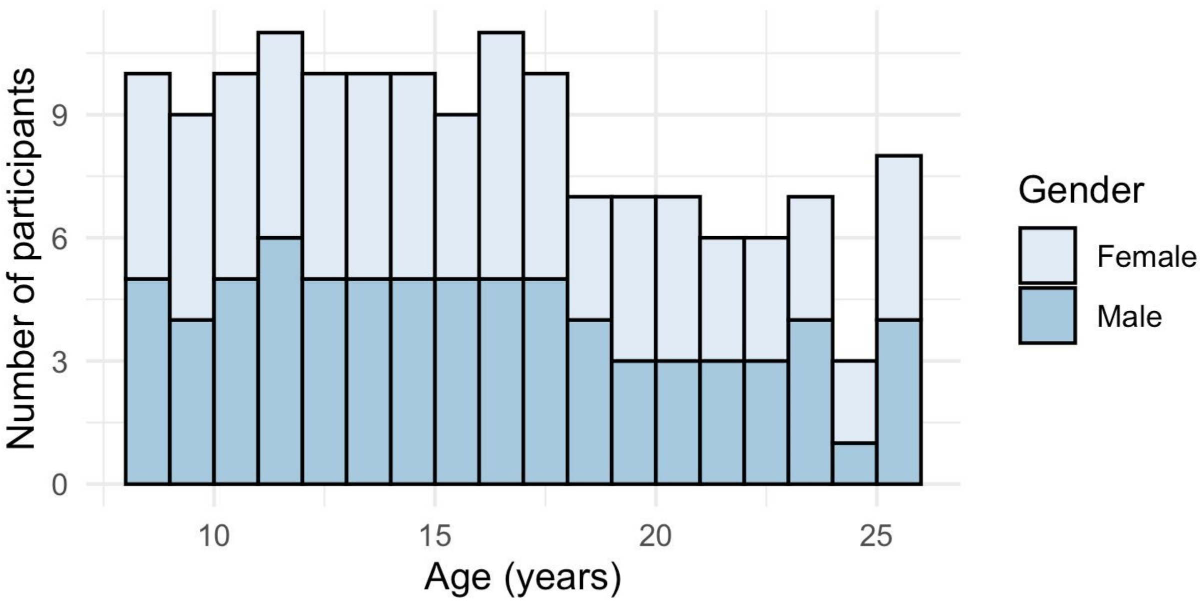

Figure S1. Distribution of online participant ages and genders.

In total, we emailed 251 potential participants (83 parents of children (8 - 12 years old), 76 parents of adolescents (13 - 17 years old), and 92 adults) with information about the study and a link to participate. Of the potential participants we emailed, 79 (22 children, 22 adolescents, and 35 adults) did not click the link prior to its expiration date, 15 (9 children, 3 adolescents, and 3 adults) filled out the consent form but did not finish the first task, and two children attempted to complete the task but could not due to technical difficulties (screen went blank partway through). In addition, we lost data from one adolescent from the two-step task, seemingly due to a technical glitch. Further, two emails to adults were returned to us due to incorrect email addresses provided at the time of sign-up.

Of the 151 participants who successfully completed both tasks, 20 (13 children, 3 adolescents, and 4 adults) accidentally quit before the automatic re-direct to the MaRS-IB task at the end of the two-step task, but all subsequently completed the MaRS-IB when we emailed them a new, direct link to the task. Eleven participants (4 children, 5 adolescents, and 2 adults) also realized on their own that they clicked the wrong button ('cancel' instead of 'leave') when asked if they wanted to load the second task (the MaRs-IB), but rather than emailing us, simply refreshed the page and did the entire two-step task again so they could proceed. In these cases, we did not analyze their second two-step task dataset. In addition, 17 participants (8 children, 3 adolescents, and 6 adults) reported that they had trouble loading the task, clicked the wrong button on the consent form, or clicked the single-use consent form when they were not ready to participate, but all of them were able to successfully load and complete the experiment when sent a new link. Finally, three parents and one adult participant emailed to say that they had not received their Amazon giftcard within three business days. Two parents were able to subsequently find the cards in their spam folder, and we canceled and re-issued giftcards for the others, who reported receiving them at that time.

Computational modeling methods and results

Hybrid reinforcement learning model. The model consists of both a "model-free" and "model-based" learning algorithm, which separately compute the value of each action (*a*) in each state (*s*) on each trial (*t*). The "model-free" algorithm computes the value of each first- and second-stage choice option (Q_{MF}) based on a standard temporal difference equation:

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

where α is a free parameter capturing individuals' learning rates, and δ is the prediction error. At the first stage, the prediction error is simply the difference between the expected value of the action they selected in the first stage and the expected value of the second-stage state-action value. At the second stage, the prediction error is the difference between the reward they actually experienced and the expected value of the action they selected in the second stage. This second-stage prediction error is then multiplied by an eligibility trace (λ), a free parameter, before it is used to update first-stage state-action values according to Equation 1, above.

This second-stage update is equivalent under the "model-based" learning algorithm. However, the first-stage learning function differs because it takes into account the 70/30 probability transition structure of the task, and computes the values of the first-stage choice based on both possible second-stage outcomes, weighted by their probabilities. The model-based algorithm assumes that the value of each second-stage state is equal to the value of selecting the highest-valued action within that state. Thus, state-action values for first-stage choices can be computed with the following equation:

$$Q_{MB}(s_{1,t}, a_{j,t}) = P(s_2|s_1, a_j) \max_{a \in \{a_A, a_B\}} Q_{TD}(s_2, a) + P(s_3|s_1, a_j) \max_{a \in \{a_A, a_B\}} Q_{TD}(s_3, a)$$
(2)

To convert these value estimates into choice probabilities, the model employs a softmax choice rule, which multiples each value estimate by separate, participant-specific inverse temperatures:

$$P(a_{i,t} = a|s_{i,t}) = \frac{exp[\beta_{MF} * Q_{MF}(s_{i,t}, a) + \beta_{MB} * Q_{MB}(s_{i,t}, a) + p * rep(a)]}{\sum_{a'} exp[\beta_{MF} * Q_{MF}(s_{i,t}, a') + \beta_{MB} * Q_{MB}(s_{i,t}, a') + p * rep(a')]} \tag{3}$$

where *p* is a free parameter that captures choice “stickiness” and rep(a) is set to 1 for repeated first-stage actions and 0 otherwise. At the second-stage, there is only a model-free Q value, and it gets multiplied by a separate, inverse temperature term (β).

Model-fitting. For each participant, we identified the parameter values that maximized the log posterior of their choices using the fmincon function in the optimization toolbox in Matlab 2020a (Mathworks, 2014). As in Decker et al. (2016) and Potter, Bryce et al. (2017), we did not include the first 9 choice trials in the analysis. We applied the following bounds and priors to each parameter: α, λ: bounds = 0, 1; prior = beta(1.1, 1.1); β_{MB}, β_{MF}, β : bounds = 0, 30; prior = gamma(3,1); *p* : bounds = -30, 30, prior = normal(0, 10). We randomly initialized parameter values, drawing uniformly from within their bounds. We initialized and ran fmincon 10 times per participant, and took the parameter estimates that maximized the log posterior across runs.

Full modeling results. The best-fitting parameters for each subject are available online on our OSF repository (https://osf.io/we89v /) . Table S1 shows mean parameter estimates for each age group in each dataset. Bolded rows indicate parameters that varied significantly as a function of age, as estimated by a linear regression. As shown in the table, results from our modeling of the online dataset and the Decker dataset aligned closely; in both cases, we observed an increase in learning rates, model-based inverse temperatures, second-stage inverse temperatures, and stickiness with increasing age.

Table S1.

Reinforcement learning model parameter estimate means (and standard deviations)

	Decker et al. (2016)			**Potter, Bryce et al. (2017)**			**Online**		
	Children	*Adolescents*	*Adults*	*Children*	*Adolescents*	*Adults*	*Children*	*Adolescents*	*Adults*
α	**.30 (.34)**	**.47 (.37)**	**.57 (.26)**	**.21 (.31)**	**.44 (.38)**	**.55 (.26)**	**.43 (.33)**	**.53 (.26)**	**.56 (.30)**
β_{MB}	**1.71 (.65)**	**2.21 (1.42)**	**2.51 (1.78)**	**2.08 (.83)**	**2.34 (1.33)**	**2.86 (1.51)**	**1.98 (.82)**	**2.86 (1.52)**	**2.86 (1.92)**
β_{MF}	2.00 (1.03)	1.84 (.97)	2.39 (1.10)	2.42 (1.61)	1.92 (.92)	1.94 (1.21)	1.88 (1.17)	2.31 (1.07)	2.25 (1.04)
β_{MB}	**1.85 (.80)**	**2.57 (1.51)**	**3.18 (1.58)**	2.89 (1.90)	2.95 (1.47)	3.08 (1.31)	**2.47 (1.52)**	**3.41 (1.56)**	**3.13 (1.45)**
λ	.51 (.28)	.51 (.33)	.61 (.30)	**.40 (.31)**	**.59 (.29)**	**.59 (.28)**	.55 (.30)	.57 (.32)	.61 (.30)
p	**.21 (.40)**	**.84 (.75)**	**1.15 (.83)**	**.11 (.58)**	**.45 (.68)**	**.99 (.72)**	**.62 (.63)**	**.97 (.83)**	**1.29 (1.23)**

Bolded rows indicate parameters that varied significantly as a function of continuous age.

Online data quality metrics

In the manuscript, we included histograms of key data quality metrics from our online dataset. Here, we report the number of subjects in each age group who met or exceeded particular quality thresholds. For the two-step task, the “fast RT” time of 150 ms was defined arbitrarily; because every trial was the same and the spaceships presented during the first-stage choice did not change sides throughout the task, participants could potentially anticipate and plan their motor response before they appeared on the screen. As such, a high number of “fast RTs” does not necessarily indicate that a participant was not engaged during the task. For the MaRs-IB, we used the same “fast RT” threshold (250 ms) as the original authors (Chierchia, Fuhrmann, et al. (2019)). Here, participants *could not* anticipate and plan their motor response prior to the trial onset.

Table S2

Number (and proportion) of participants meeting data quality thresholds in the two-step task

Metric		**Children**	**Adolescents**	**Adults**
Browser Interactions	≤ 3	39 (.78)	41 (.82)	41 (.80)
	≤ 5	45 (.90)	46 (.92)	43 (.84)
	≤ 10	48 (.96)	46 (.92)	46 (.90)
Comprehension Questions Correct	≥ 1	50 (1)	50 (1)	51 (1)
	≥ 2	49 (.98)	50 (1)	50 (.98)
	3	44 (.88)	40 (.80)	45 (.88)
Missed Responses	≤ 10 (2.5%)	39 (.78)	48 (.96)	45 (.88)
	≤ 20 (5%)	44 (.88)	49 (.98)	48 (.94)
	≤ 40 (10%)	49 (.98)	49 (.98)	49 (.96)
RTs < 150 ms	≤ 10 (2.5%)	17 (.34)	30 (.60)	25 (.49)
	≤ 20 (5%)	27 (.54)	38 (.76)	31 (.61)
	≤ 40 (10%)	37 (.74)	43 (.86)	48 (.94)

Table S3

Number (and proportion) of participants meeting data quality thresholds in the MaRs-IB task

Metric		**Children**	**Adolescents**	**Adults**
Browser Interactions	≤ 3	47 (.94)	45 (.90)	47 (.92)
	≤ 5	49 (.98)	47 (.94)	49 (.96)
	≤ 10	50 (1)	50 (1)	51 (1)
Practice Trials Needed	3	28 (.56)	41 (.82)	41 (.80)
	≤ 4	39 (.78)	48 (.96)	50 (.98)
	≤ 5	45 (.90)	48 (.96)	51 (1)
Missed Responses	≤ 3	49 (.98)	50 (1)	51 (1)
	≤ 5	50 (1)	50 (1)	51 (1)
RTs < 250 ms	≤ 3	45 (.90)	49 (.98)	47 (.42)
	≤ 5	45 (.90)	49 (.98)	48 (.94)
	≤ 10	47 (.94)	50 (1)	48 (.94)