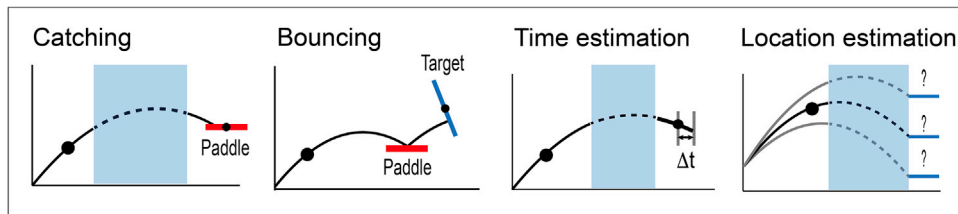


Article

Developmental change in predictive motor abilities

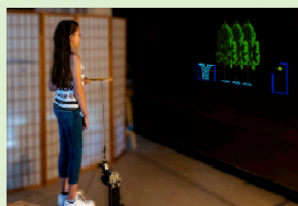
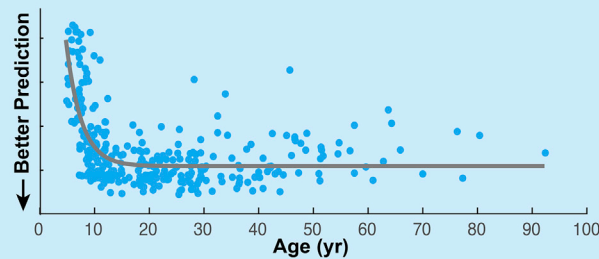
Development of predictive motor abilities



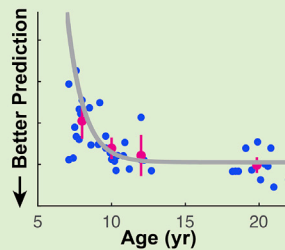
Data from four tasks across two settings showed that by 12 years of age, predictive abilities become indistinguishable from adult performance.



Museum Setting



Laboratory Setting



Se-Woong Park,
Annie Cardinaux,
Dena Crozier,
Marta Russo,
Margaret
Kjelgaard, Pawan
Sinha, Dagmar
Sternad

sewoong.park@utsa.edu

Highlights

Development of
predictive motor ability
was tested in 4 virtual
interception tasks

Among 442 subjects aged
5–92, predictive abilities
reached adult level by age
12

Findings were consistent
between public museum
and controlled laboratory
settings

Provides normative data
for development of
prediction in conditions
like autism

Park et al., iScience 26, 106038
February 17, 2023 © 2023 The
Author(s).
<https://doi.org/10.1016/j.isci.2023.106038>

Article

Developmental change in predictive motor abilities

Se-Woong Park,^{1,2,3,10,*} Annie Cardinaux,³ Dena Crozier,^{2,4,5} Marta Russo,^{6,7} Margaret Kjølgaard,⁸ Pawan Sinha,³ and Dagmar Sternad^{2,4,9}

SUMMARY

Prediction is critical for successful interactions with a dynamic environment. To test the development of predictive processes over the life span, we designed a suite of interceptive tasks implemented as interactive video games. Four tasks involving interactions with a flying ball with titrated challenge quantified spatio-temporal aspects of prediction. For comparison, reaction time was assessed in a matching task. The experiments were conducted in a museum, where over 400 visitors across all ages participated, and in a laboratory with a focused age group. Results consistently showed that predictive ability improved with age to reach adult level by age 12. In contrast, reaction time continued to decrease into late adolescence. Inter-task correlations revealed that the tasks tested different aspects of predictive processes. This developmental progression complements recent findings on cerebellar and cortical maturation. Additionally, these results can serve as normative data to study predictive processes in individuals with neurodevelopmental conditions.

INTRODUCTION

A soccer goalkeeper blocking a penalty shot, a car driver navigating traffic, and a child playing a game of catch—all these actions have one key commonality: the actor must predict the behavior of the moving object. In order to succeed at these tasks, a purely reactive stance is inadequate; the temporal sluggishness of the perceptuo-motor loop renders *re*-actions alone ineffectual. Predictions about an event that is yet to transpire are required to successfully perform these interceptive actions.^{1–3} But the relevance of prediction is not restricted to interactions with external objects. Almost any coordinated action relies on prediction. For example, while standing, postural adjustments are required prior to performing arm actions to avoid losing our postural balance.⁴ Any goal-directed movement requires prediction of the sensory consequences of one's actions to recognize control errors and distinguish between self and others' errors.^{2,5} When manipulating objects with additional dynamics, such as a cup filled with coffee, prediction-based feedforward control has been shown to be paramount as error correction based on sensory feedback is too slow.^{6–8}

The critical role of prediction not only in motor behavior, but also in perception and cognition has long been recognized and is currently the focus of much research.^{9,10} The framework of predictive coding has been developed to understand computational and neural mechanisms underlying cognitive, language, and perceptual processes.^{11,12} To account for the inherent uncertainty in these processes, Bayesian theory has been leveraged where inferences based on sensory information and prior experiences afford predictions to guide decisions about future events.^{13,14} While numerous studies detailed potential computational and neural mechanisms for cognition, speech, and perception, comparatively little research has turned to the motor domain. This study examines predictive proficiency and its developmental progression in the motor domain.¹⁵ We use interceptive actions with an external object to empirically probe predictive abilities from childhood into adulthood.

Behavioral assessments of motor development have focused on the emergence and improvement of coordinative milestones, such as crawling, sitting, and walking, based on behavioral observations with standardized administration and scoring that may be subject to individual clinicians' interpretation.^{16,17} Many experimental investigations of motor skill development have been largely descriptive and confined to performance measures, such as error or the number of successful executions.^{18,19} The development of interceptive skills and specifically catching has been frequently studied but primarily with qualitative and often inconsistent approaches.^{20–23}

¹Department of Kinesiology, University of Texas at San Antonio, San Antonio, TX 78249, USA

²Department of Biology, Northeastern University, Boston, MA 02115, USA

³Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁴Department of Physics, Northeastern University, Boston, MA 02115, USA

⁵Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA

⁶Department of Neurology, Tor Vergata Polyclinic, Rome, Italy

⁷Laboratory of Neuromotor Physiology, IRCCS Santa Lucia Foundation, Rome, Italy

⁸Department of Communication Sciences and Disorders, Bridgewater State University, Bridgewater, MA 02325, USA

⁹Department of Electrical & Computer Engineering, Northeastern University, Boston, MA 02115, USA

¹⁰Lead contact

*Correspondence:

sewoong.park@utsa.edu

<https://doi.org/10.1016/j.isci.2023.106038>



Relatively few studies have employed quantitative methods from computational motor control to understand the developmental progression of sensorimotor skills. For example, some studies examined trajectory planning in younger children using planar reaching to targets as experimental testbed.^{24–27} Detailing after-effects to visuomotor rotations, the work identified age-dependent formation of internal representations and their visual and kinesthetic reliance. Implicit skill learning over the life span has also been examined in serial reaction time tasks that measured reaction times to high- and low-frequency events. Results reflect different learning processes that either rely on probabilistic model-free learning versus model-based learning or online trial-by-trial and offline processes.^{28,29} However, further research is needed to characterize the relationship between implicit skill learning and predictive processes.

Hence, the existing data do not enable quantitative characterization of the development of specifically predictive abilities in movement coordination. We present a novel test battery that aims to quantify the development of predictive abilities in the context of sensorimotor coordination, using game-like interception tasks that are easy to deploy across a broad age spectrum. While this study focuses on motor development in young children, our data span the age range from 5 to 92 years.

Characterizing this developmental progression is relevant not only for understanding maturation in typical development, but also in conditions that might diverge from this trajectory. Several neurological and neurodevelopmental conditions are associated with sensorimotor coordination challenges that might implicate predictive challenges.^{30–32} For example, individuals with autism spectrum disorder (ASD) often exhibit reduced motor proficiency compared to their age-matched peers, although this aspect of the phenotype is not referenced in the DSM-5 criteria for autism.^{33–35} There is growing evidence that the atypical motor characteristics in autism may arise from a more general difference in predictive abilities that affect multiple domains, including social interactions, language learning, and sensorimotor actions.^{36–38} Investigating the development of these abilities is important not only for understanding the genesis of motor challenges associated with autism, but also for devising tests to improve early diagnosis. Addressing these questions about atypical predictive motor skills first requires characterization of the typical developmental arc of this domain.

We designed a suite of four interactive visuomotor tasks that present predictive challenges with titrated levels of difficulty to quantify predictive abilities in different scenarios. The tasks involve a virtual ball launched to follow a ballistic trajectory. Two tasks require the participant to move a paddle to catch or bounce a ball, and two require only button presses to indicate when or where a ball will appear. An additional reaction time task serves as a comparison to quantify basic processing speed. [Figure 1](#) depicts the five screen-based tasks.

The “catching” task required interception of a ball with a hand-held paddle that was represented as a virtual basket on the screen. To succeed, a participant needed to predict the ball location unfolding in time *in relation* to the paddle. To vary the difficulty of prediction, the ball trajectory was partly covered by an occlusion of three different sizes. The “bouncing” task required the same paddle movements to not only intercept, but also bounce the ball to a target. Therefore, predicting the ball’s trajectory and achieving the right *position and velocity* at contact was necessary to reach the target. To reduce motor dexterity demands, two additional tasks (“pausing” and “choosing”) employed simple key presses, rather than paddle manipulation. In the “pausing” task, participants had to predict *when* the virtual ball would appear in a target window by pressing a button to stop the ball’s movement. Due to occlusion, the ballistic ball trajectory had to be extrapolated to correctly estimate the time of reappearance in the window. In the “choosing” task, participants had to predict *where* the ball would emerge following occlusion, by pressing one of three buttons representing discrete locations on the screen; the response did not require a temporally accurate key press. The “reacting” task required a fast movement of the paddle device upon a randomly timed appearance of the static target stimulus on the screen; no prediction was needed. The action involved the same paddle movement as in the interception tasks. Detailed descriptions are provided in [STAR Methods](#).

We collected data for all five tasks in two settings: the public “Living Laboratory” at the Museum of Science in Boston and a conventional laboratory. Our intent in using the two sites was to test a broader sample of participants to gain first insights, and then narrow the focus to test the relevant age range in the laboratory

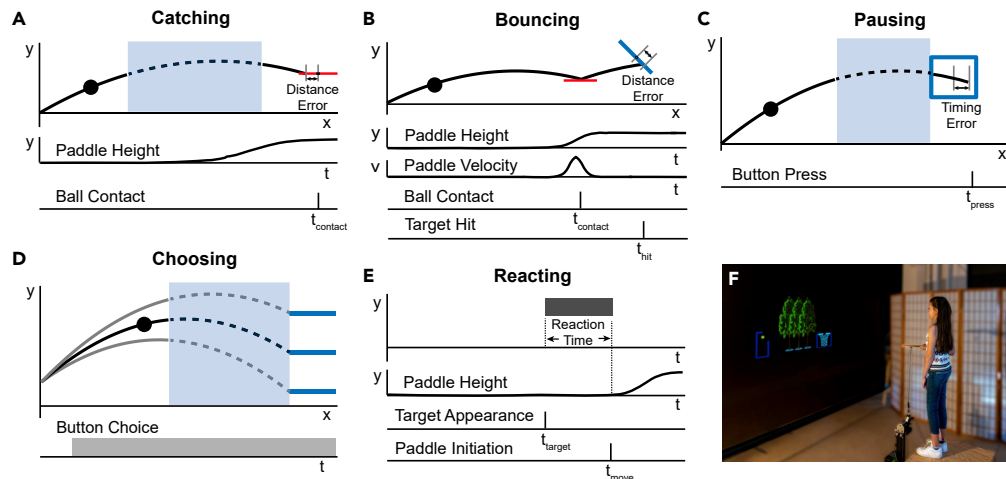


Figure 1. Design of five virtual tasks

(A) **Catching**: A ball is launched on one of two different trajectories. The blue area indicates partial occlusion of the ball trajectory to increase the predictive challenge inherent in this task. The participant is asked to catch the ball, i.e., predict the best time and location of the ball for interception at time t_{contact} . Performance is quantified by the catch error, defined as the distance between the center of the paddle (red line) and the contact location of the ball. The result is determined by the paddle height y at t_{contact} .

(B) **Bouncing**: The task requires participants to intercept and bounce the ball to a target (slanted blue line). Unlike in catching, the result is determined not only by paddle height but also by paddle velocity at time t_{contact} . Performance is quantified by the distance between the ball contact location on the target and the center of the target.

(C) **Pausing**: The task requires participants to press a button at the time when the ball reappears in the center of a window (blue square). The blue shaded area occludes the ball after launch to elevate demands on prediction. Performance is quantified by the time between the button press (t_{press}) and the time when the ball is in the center of the window.

(D) **Choosing**: The task requires participants to predict the landing location of the ball by pressing one of three buttons. There is no requirements on timing (gray-shaded bar). Performance is categorical indicating the correct or incorrect landing location.

(E) **Reacting**: A static object stimulus appears at a randomized time on the screen. The task is to lift the paddle as fast as possible upon the object appearance at t_{target} . Performance is quantified by the time between t_{target} and initiation of the paddle movement t_{move} .

(F) A participant in the laboratory's virtual environment holds the paddle while awaiting the catching task ball launch.

with in-depth analyses. The museum setting also allowed us to enroll a large number of individuals (a total of 442 participants) across a wide age range (5–92 years). Across this large group, we hoped to discern the overall trends of developmental change, although we expected that the data would have significant variability due to the presence of extraneous factors in the public space. Laboratory-based studies were designed to complement the museum-based investigations by allowing longer testing times, a more controlled environment, and more selective enrollment with regard to age groups and specific inclusion/exclusion criteria. A welcome benefit of working in the museum was social outreach—informing the public about the research process, and encouraging their participation in scientific studies, while also articulating their relevance to real-world issues.

RESULTS

We first examined the performance changes across ages with the help of an exponential regression to explore age-related changes across the life span. Subsequent analyses aimed to identify at what age children's performance approached that of adults. To this end, young participants were parsed into six age groups each spanning two years (5–6, 7–8, 9–10, 11–12, 13–14, and 15–16 years old) to afford comparison with the young adult group (ages 18–21 years). Note that in the museum, the number of participants varied across the 5 games due to the different amount of time museum visitors had been available to engage in our studies, ($n = 288, 286, 322, 317, \text{ and } 269$). For the laboratory data, we performed a similar analysis to complement and verify the results gained at the museum, testing participants in a more controlled environment. Forty-four participants (none of whom had participated in the museum-based studies) were divided into four age bins, of consistent width as in the museum data (ages 7–8, 9–10, 11–12, and 18–21 years old).

Table 1. Comparisons between children age groups with the young adult group (18–21 years) at the museum

	Museum					
Age (years)	5–6	7–8	9–10	11–12	13–14	15–16
Catching	$p < 0.001$ $\beta = 39.604$	$p < 0.001$ $\beta = 27.465$	$p < 0.001$ $\beta = 20.567$	$p < 0.001$ $\beta = 16.857$	$p = 0.012$ $\beta = 17.257$	$p=0.492$ $\beta =5.154$
Bouncing	$p < 0.001$ $\beta = 83.540$	$p < 0.001$ $\beta = 47.603$	$p < 0.001$ $\beta = 41.516$	$p < 0.01$ $\beta = 33.092$	$p=0.598$ $\beta =9.856$	$p=0.847$ $\beta =-3.969$
Pausing	$p < 0.001$ $\beta = 86.380$	$p < 0.001$ $\beta = 47.086$	$p < 0.001$ $\beta = 21.535$	$p < 0.01$ $\beta = 18.265$	$p=0.816$ $\beta =2.140$	$p=0.161$ $\beta =17.186$
Choosing	$p < 0.001$ $\beta = -1.570$	$p < 0.001$ $\beta = -0.852$	$p < 0.001$ $\beta = -0.663$	$p=0.246$ $\beta =-0.231$	$p=0.895$ $\beta =-0.048$	$p=0.672$ $\beta =0.156$
Reacting	$p < 0.001$ $\beta = 92.461$	$p < 0.001$ $\beta = 45.920$	$p < 0.001$ $\beta = 30.930$	$p < 0.01$ $\beta = 20.381$	$p=0.705$ $\beta =-4.284$	$p = 0.014$ $\beta = 31.136$

GLMM results from comparisons between each age group with the adult group. β is the coefficient for the age effect. The bolded numbers indicate that the comparison with the adult group was not significant, marking that performance in this age group was not distinguishable from adults.

These age groups covered the most relevant age range for our question: at what age does predictive ability reach adult performance.

Results were obtained by using a generalized linear mixed model (GLMM) that included all trials to account for inter- and intra-individual variability (see [STAR Methods](#)). The complete set of statistical results is summarized in [Tables 1](#) and [2](#). We also examined cross-task correlations to assess independence of results and potential redundancies in our test battery.

Task 1: Catching

Catching the ball in this task requires moving the paddle to the right place at the right time. The experimental task included paddle interception of the ball after partial occlusion, represented as one to three trees on the screen ([Figures 2A](#) and [2B](#); for more details see [STAR Methods](#) and [Figure 8](#)). The experiment measured paddle-ball interception accuracy by quantifying how close the ball was to the center of the paddle at contact, i.e., the error served as a performance metric. At the museum, participants performed only 15 trials in this task, due to the time constraints set by museum policy (a maximum overall experiment session duration of 15 min). In the laboratory, participants performed 60 trials, with 20 trials per occlusion condition. The absolute error was calculated for each trial of each participant.

[Figure 2C](#) summarizes catching performance for all museum participants ($n = 288$). Visual inspection of all the participants shows that the absolute error decreased with age, i.e., performance improved. The GLMM analysis confirmed this, using age as a continuous factor ($\beta = -0.53$, $p < 0.001$). More specifically, performance error dropped until adolescence and then plateaued during adulthood. To focus on the developmental trajectory in childhood and adolescence (5–21 years), participants were grouped into 2-year age bins and averaged, as shown by the red points in [Figure 2D](#). Children exhibited decreasing error as age increased, although with high inter-individual variability. While the differences between younger groups and the adult group were significant, this difference disappeared at age 15–16 ($\beta = 5.154$, $p = 0.492$), suggesting that predictive interception ability at age 15–16 approached that of the adults. These and all other museum-based results are summarized in [Table 1](#).

To further scrutinize these results, we analyzed the laboratory data to discern the age at which the performance became indistinguishable from the adults ([Figure 2E](#)). The older children expectedly outperformed the younger children. In this more controlled setting, the group of 11- to 12-year-old children already achieved performance that was statistically indistinguishable from the adult group ($\beta = 9.80$, $p = 0.102$). Details of these and all other laboratory results are summarized in [Table 2](#).

Effect of occlusion

The performance asymptote shown with age was also evident when the data were split by the degree of occlusion (number of trees). In the museum data, the GLMM also revealed an occlusion effect, i.e., the performance

Table 2. Comparisons between each age group with the young adult group (18–21 years) in the laboratory

Age (years)	Laboratory		
	7–8	9–10	11–12
Catching	$p < 0.001$ $\beta = 38.781$	$p < 0.001$ $\beta = 20.427$	$p=0.102$ $\beta =9.797$
Bouncing	$p < 0.001$ $\beta = 64.867$	$p = 0.050$ $\beta = 23.057$	$p=0.232$ $\beta =17.140$
Pausing	$p < 0.001$ $\beta = 38.718$	$p < 0.01$ $\beta = 20.830$	$p=0.401$ $\beta =6.448$
Choosing	$p < 0.001$ $\beta = -1.447$	$p < 0.001$ $\beta = -1.308$	$p=0.669$ $\beta =0.194$
Reacting	$p < 0.001$ $\beta = 79.193$	$p < 0.001$ $\beta = 56.092$	$p = 0.035$ $\beta = 23.514$

GLMM results from comparisons between each age group with the adult group. β is the coefficient for the age effect. The bolded numbers indicate that the comparison with the adult group was no longer significant, indicating that this age group was no longer distinguishable from adult performance.

depended on the amount of occlusion ($\beta = 4.06$, $p < 0.001$). Albeit significant, the effect sizes were small across the age ranges (inset of Figure 2C), as also confirmed by the effect in the two-year age bins analysis ($\beta = 3.34$, $p < 0.01$). Similarly, the laboratory data were divided into the same three occlusion (tree) conditions. However, the results did not reveal an age-related dependency for occlusion effects ($\beta = 0.50$, $p = 0.62$).

Task 2: Bouncing

Distinct from catching a ball, bouncing the ball to a target required contacting the ball not only at the right *location* but also at the right *velocity* to appropriately direct the subsequent flight path. Performance error was quantified as the distance between where the ball crossed the dashed target line and the line's center shown as a red dot (Figure 3A; see also Figure S2, illustrating how the target was represented as a crocodile's open mouth to make the task engaging for children). At the museum, participants performed 15 trials, whereas participants who visited the laboratory performed 60 trials.

Figure 3B displays the bouncing task performance of all museum participants ($n = 286$) with an exponential function fit over the entire age range. As in the catching game, a GLMM revealed a main effect of age ($\beta = -1.09$, $p < 0.001$). As is evident in the exponential fit, younger participants displayed worse performance than older participants. Figure 3C shows the museum results comparing children to adults (age 18–21). Children exhibited a decrease in error with high variability as age increased; performance at age 13–14 reached that of adults ($\beta = 9.86$, $p = 0.598$) and age 15–16 ($\beta = -3.97$, $p = 0.847$) showed the same trend (Table 1).

Figure 3D displays the laboratory results. While bouncing performance at age 7–8 was significantly worse than in the adult group ($\beta = 64.87$, $p < 0.001$), the performance improved at age 9–10 ($\beta = 23.06$, $p = 0.05$), and became comparable to adults by age 11–12 ($\beta = 17.14$, $p = 0.23$; see Table 2).

Task 3: Pausing

While the previous games involved arm-hand coordination, the next two games reduced motor involvement to only a key press. First, pausing task measured the error in time as participants tried to stop a ball after its launch, disappearance behind occlusion (house), and reappearance in a window. Using the center of the window as reference, the error was the difference between the time when the ball was stopped and the time it would have reached the center of the window. Figure 4A shows how the pausing task was operationalized for participants. The trajectory of the launched ball was partially occluded by entering a house and the participant's task was to pause the ball when it re-appeared in the crosshairs of a window. A purely reactive strategy, pressing the key after the ball re-appeared in the window, would fail since the ball would move past the window by the time of the response. At both the museum and the laboratory, participants performed 30 trials of the pausing task.

Figure 4B shows the temporal prediction results of all museum participants across all ages ($n = 322$) and their fit with an exponential function. Consistent with the catching and bouncing tasks, a GLMM revealed that children

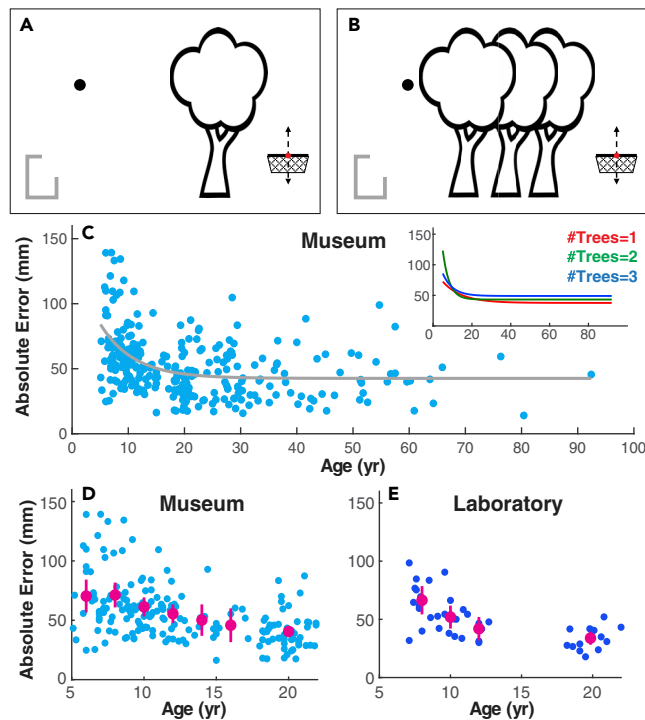


Figure 2. Catching task

(A and B) Schematic of the task display. A ball is launched from the box on the left side of the screen; one to three trees occlude the ball trajectory. Participants were instructed to lift the basket on the right to intercept the ball at the center of the basket (shown by a red dot).

(C) Individual median absolute catch-errors across the full museum sample ($n = 288$), spanning an age range of 5–92 years. The gray line shows the exponential fit ($y = ae^{-bt} + c$) for the entire dataset. In the inset, the red, green, and blue lines show the exponential fits when the errors were analyzed separately for trials with one, two, and three trees, respectively.

(D) Catch-errors for the museum experiment as a function of age when limited to 21 years.

(E) Catch-errors as a function of age in the laboratory experiment. The red points are the means per age bin, the vertical lines indicate the 95% confidence interval for each age bin.

performed with significantly higher errors than the adults ($\beta = -0.61$, $p < 0.001$); the errors approached the asymptote around 12–15 years. To detail this observation, Figure 4C shows the museum data focusing on the relevant age range with children grouped into two-year bins. Participants showed a clear decline in their temporal errors with increasing age and the GLMM revealed that the 13–14 and the 15–16-year-old groups were statistically indistinguishable from young adults ($\beta = 2.14$, $p = 0.816$; $\beta = 17.18$, $p = 0.161$, respectively). The laboratory data, shown in Figure 4D, revealed that even the 11–12-year-olds had approached performance of the adult group (18–21 years, $\beta = 6.45$, $p = 0.401$).

Effect of ball speed

To determine whether the better performance of the older age group (11–12-year-olds) relative to the younger one (7–8-year-olds) was due to greater variability in the latter, we examined whether the decreasing timing error was related to an increased sensitivity to ball speed. Figures 4E and 4F show the laboratory data of all timing errors for the three ball speeds (or target times) of one representative child from each of the two age groups. The younger child pressed the button with high variability and the mean timing did not vary with target time. In contrast, the older child performed with less variability and the mean performed time increased with longer ball flight times with a highly significant linear progression that mirrored the target times (the dotted line shows the identity line for target time and performed time). To quantify sensitivity to the target time for children from the museum and laboratory, the slopes of the linear regressions for all participants were computed and summarized in Figure 4G; the red points show the mean values of the slopes for the four age groups. The mean slope values indicate that the sensitivity to ball speed increased with age, approaching a slope of 1, corresponding to perfect temporal prediction.

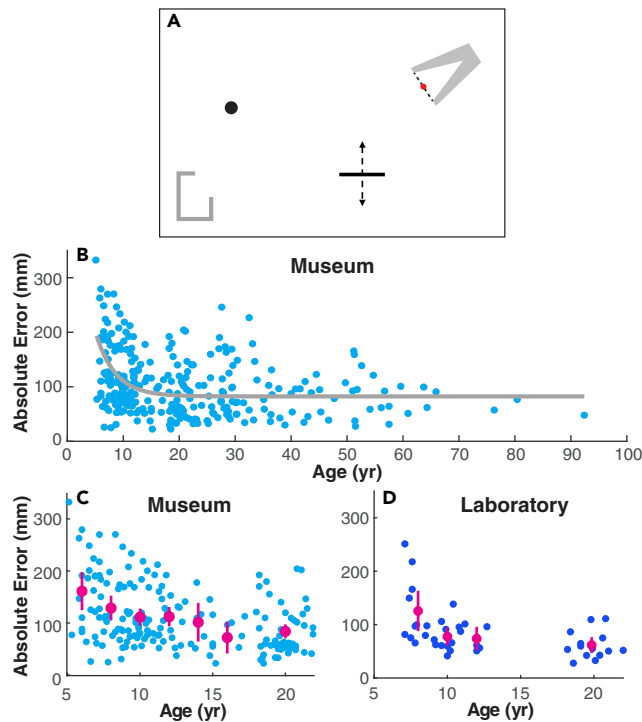


Figure 3. Bouncing task

(A) Schematic of the task display. The instruction was to bounce the approaching ball to hit the center of the open ‘mouth’ on the right (shown by a red dot in the center of the dashed line).

(B) Individual median absolute target errors across the full museum sample ($n = 286$), spanning an age range from 5 to 92 years. The gray line is the exponential regression fit ($y = ae^{-bt} + c$).

(C) Target errors as a function of age groups in the museum experiment. The red points are the means per age bin; the vertical lines indicate the 95% confidence interval for each age bin.

(D) Target errors as a function of age for the age range from 7 to 12 in the laboratory. The red points are the means per age bin; the vertical lines indicate the 95% confidence interval for each age bin.

Task 4: Choosing

This task required participants to make a spatial prediction about where a ball will land, based on a partially observed trajectory (Figure 5A). The metric for this task was not continuous because the response options were limited to one of three discrete landing locations. We therefore evaluated performance as trial success (0 or 1) for selecting the correct location. The figures present performance as success rate defined as the percentage of successful trials. At both the museum and the laboratory, participants performed 30 trials of the choosing task.

Figure 5B shows increasing success rate for all museum participants ($n = 317$) with age, again highlighted by an exponential regression fit. The GLMM analysis confirmed the significant effect of age ($\beta = 0.07$, $p < 0.001$). Figures 5C and 5D show participant success rates with a focus on the younger groups, for the museum and the laboratory data. In both settings, performance of the group of 11-12-years-old children became statistically indistinguishable from that of adults (museum: $\beta = -0.23$, $p = 0.246$; laboratory: $\beta = 0.19$, $p = 0.669$; see also Tables 1 and 2).

Effect of location

Closer examination of the data revealed that the success rates differed for the three landing locations. The museum and laboratory data in Figure 5E and F show a split for the three target locations, indicating that the red target had higher success rates than the two lower targets (green and blue). This difference was confirmed by a significant effect in the GLMM in both museum and laboratory data (museum: $\beta = -0.29$, $p < 0.001$; laboratory: $\beta = -0.48$, $p < 0.001$). The three schematic ball trajectories displayed in Figure 5A provide some insight into these differences. The blue (bottom) trajectory just prior to occlusion had the greatest

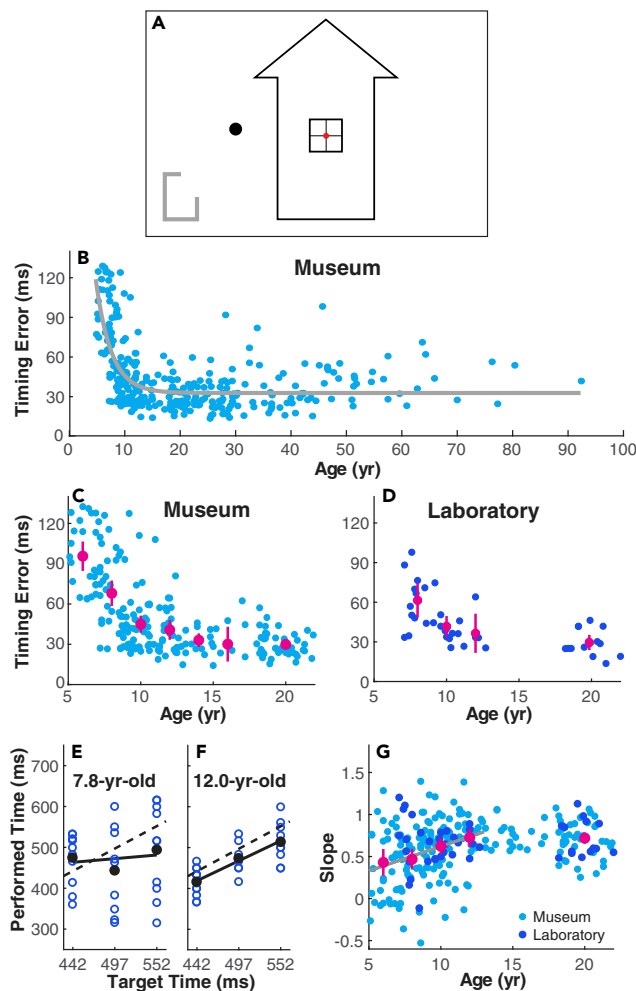


Figure 4. Pausing task

(A) Schematic of the task display. Participants saw the initial trajectory of the ball after launch, which was then occluded by a house. They were required to press a key to pause the ball when it re-emerged in the center of the window.

(B) Individual median temporal errors across the full museum sample ($n = 322$), spanning an age range of 5–92 years. The gray line is the exponential regression fit ($y = ae^{-bt} + c$).

(C and D) Temporal errors as a function of age in the museum and laboratory experiment. The red points are the means per age bin; the vertical lines indicate the 95% confidence interval for each age bin.

(E and F) Time of button press for two example children at the laboratory for the three ball-flight speeds (defining the target time). The black dotted identity line represents perfect prediction. The 7.8-year-old child pressed the button with high variability and the mean time did not differ for the three ball trajectories. The older child's press times show a systematic increase as a function of ball flight times.

(G) Regression slopes performed as shown in panels E and F of all participants as a function of age in the laboratory and the corresponding ages in the museum group. The gray line shows a linear regression of the slopes across ages 5 to 12.

offset to the correct (bottom blue) target compared to the red and green trajectories. This offset appeared to present difficulties for participants since performance with the bottom target was consistently lower than for the other two.

Task 5: Reacting

As a control, we also assessed reaction time in a custom-developed task using the same child-friendly theme and the same arm/paddle movement as the catching and bouncing tasks (Figure 6A). Participants were instructed to lift the paddle (depicted as a basket) to catch a “mouse” as fast as possible after its appearance. Reaction time was computed as the time difference between the appearance of the mouse

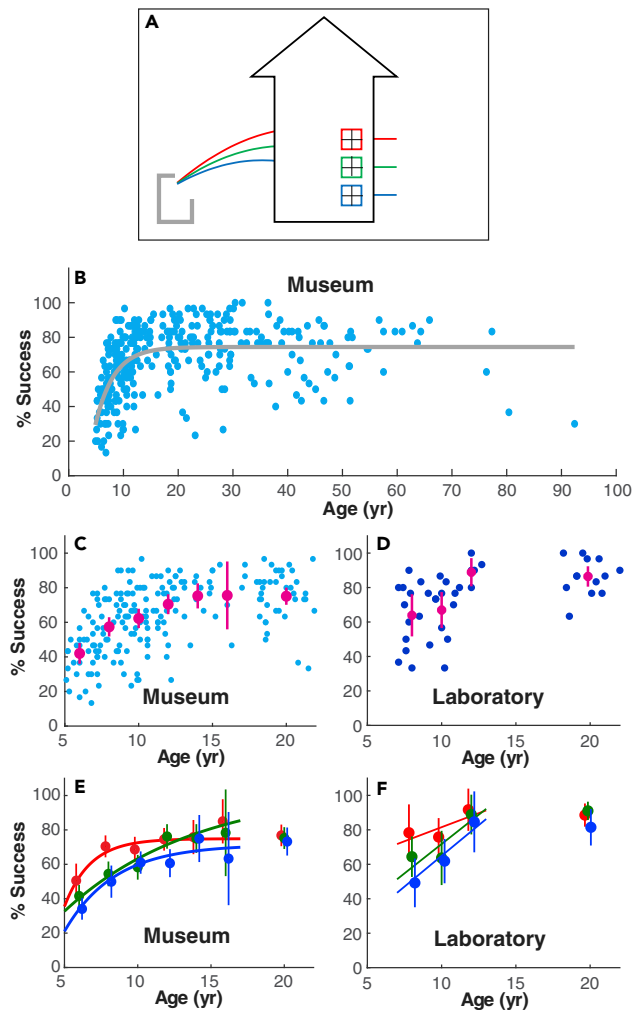


Figure 5. Choosing task

(A) Schematic of task display. Participants saw one of three ball trajectories, which was then occluded by a house. The ball would land at one of three locations on the right side of the house. The goal was to select the correct destination window by pressing one of three keys.

(B) Success rates across the full museum sample ($n = 317$), spanning an age range of 5–92 years. The gray line represents the exponential regression fit ($y = ae^{-bt} + c$).

(C) Success rates as a function of age in the museum experiment. The red points are the means per age bin, the vertical lines indicate the 95% confidence interval for each age bin.

(D) Success rates as a function of age in the laboratory experiment. The red points are the means per age bin, the vertical lines indicate the 95% confidence interval for each age bin.

(E and F) Same data as in C and D after splitting the trials by the ball destination. Red denotes the top window, green the middle window, and blue the bottom window in the house.

and the onset of the paddle movement. At the museum, participants performed 10 trials, whereas participants in the laboratory performed 20 trials.

Figure 6B summarizes the reaction times of all museum participants ($n = 269$). The data revealed a clear decline (improving performance) across age ($\beta = -0.81$, $p < 0.001$), though the leveling out of the asymptote with age was slightly less marked in comparison to previous measures. Figure 6C limits the age range to the younger groups and shows that younger children were significantly outperformed by older children: the mean reaction time at age 7–8 was 282 ms, which shortened to 257 ms at age 11–12. Participants of ages 13–14 reached the speed of reaction of adults (mean = 238 ms, $\beta = -4.28$, $p = 0.70$). However, the reaction times for ages 15–16 were significantly slower than those of the adults

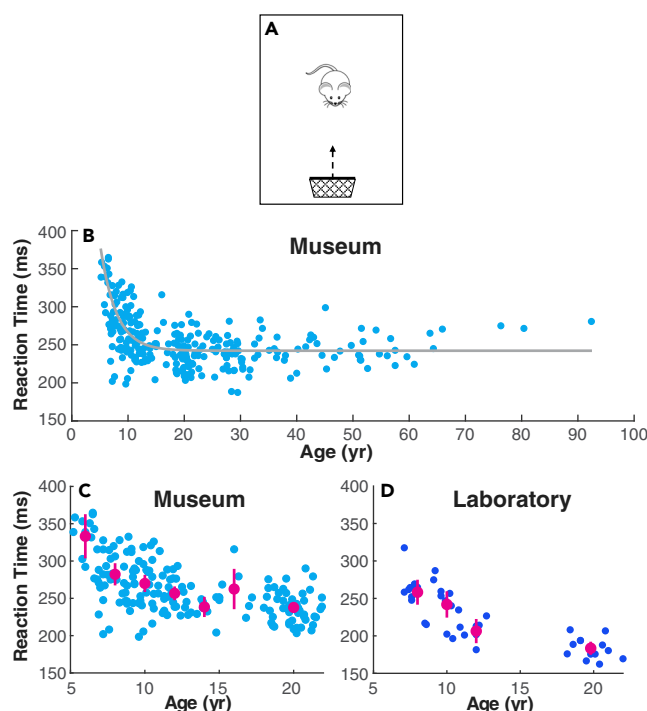


Figure 6. Reacting task

(A) Schematic of task display. Upon randomly timed appearance of a static mouse, participants were instructed to “catch the mouse” as fast as possible. Reaction time was calculated as the difference between appearance of the mouse and commencement of paddle movement.

(B) Individual median reaction times for all museum participants ($n = 269$) across ages 5 to 92 years. The gray line represents the exponential regression fit ($y = ae^{-bt} + c$).

(C) Reaction times from the museum experiment. The red points are the means per age bin, the vertical lines indicate the 95% confidence interval for each age bin.

(D) Reaction times from the laboratory experiment. The red points are the means per age bin, the vertical lines indicate the 95% confidence interval for each age bin.

($\beta = 31.13$, $p < 0.05$), possibly driven by the relatively small sample size. [Tables 1](#) and [2](#) summarize these results.

Reaction times in the laboratory overall were faster than those in the museum, likely due to the controlled environment that enhanced concentration on the games. Younger children were slower than older children ([Figure 6D](#)): the mean reaction time at age 7–8 was 258 ms and decreased to 206 ms at age 11–12; the adult mean was 183 ms. Unlike in the aforementioned prediction tasks, even at age 11–12 the children’s reaction times had not yet reached those of adults ($\beta = 23.51$, $p = 0.03$), suggesting continued development even after the age of 12, as also shown in the museum data.

Summary of results

Participants showed largely consistent age-related changes across the four prediction tasks in the museum and in the laboratory. Overall, we observed better performance in the more controlled laboratory setting, systematically shifting the transition from developing to adult-like performance to a slightly lower age than at the museum. In the museum, participants aged between 11 and 16 began to show adult-like behavior, while in the laboratory this equivalence was already evident in the 11- to 12-year-old children. Reaction time continued to improve in the laboratory even past the age of 12, and in the museum data adolescent performance appeared unstable due to the relatively small sample size. [Tables 1](#) and [2](#) summarize the statistical results.

Inter-task correlations

Pairwise comparisons between the five tasks were conducted to examine consistencies or redundancies across the five tests. These analyses only included participants who completed all five games at the

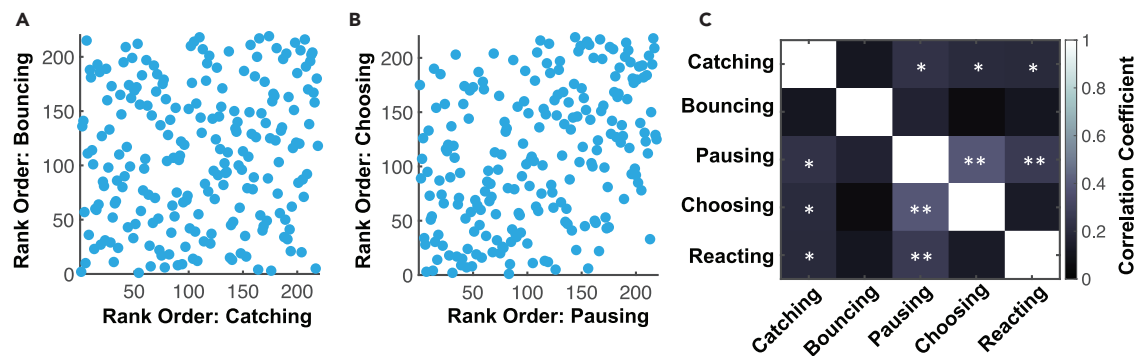


Figure 7. Inter-task correlations

(A) Correlation of individuals performance between the catching and bouncing tasks in the museum (rank 1 is best performance).

(B) Correlation of the performance between the pausing and choosing tasks in the museum.

(C) All pairwise inter-task comparisons with correlation values indicated by grayscale. The asterisks indicate significant inter-task correlation pairs (*p < 0.05, **p < 0.005, corrected). Due to the large number of subjects, even the significant correlation between pausing and choosing is only modest, as seen in (B).

museum (n = 219). To mitigate confounding by age-related variability, we implemented partial correlations with age as the controlling variable (see [STAR Methods](#)). To ensure reasonable matching, the participants were rank-ordered by their performance before computing the correlations. [Figure 7](#) reveals that all pairwise comparisons exhibited at most a weak correlation with Spearman correlation coefficients of less than 0.4. Even though the pausing-choosing correlation reached significance, probably due to the large number of subjects, the data distributions shows only a very modest correlation.

DISCUSSION

The present study investigated the development of predictive skills in the motor domain, specifically when interacting with dynamic objects, in children and adults. We used custom-developed virtual testbeds that quantified different aspects of predictive abilities: dynamic prediction with interception and control of the object using a hand-held paddle, and temporal and spatial prediction using button presses to simplify motor demands. Reaction time was tested for reference. Our suite of games allowed for the quantification and titration of predictive motor skills not only in the laboratory, but also in a public space where we could enlist a large and diverse cohort across a broad age range and test the sensitivity of our novel games. The consistent finding across tasks was that predictive performance improved significantly until the age of 11–12 years when results became comparable to those of adults. The transition to adult level occurred slightly later in the museum data which is likely due to reduced attentional focus in the less controlled environment. Importantly, this evolution of predictive skill was independent of motor dexterity demands as both paddle interactions and simple button presses exhibited similar time courses.

By contrast, reaction times revealed continued improvements into later adolescence. This difference between prediction and reaction time indicates the sensitivity of our metrics, providing indirect support that the rise and plateau in predictive skills is not merely attributable to changes in basic processing speed. It is noteworthy that these results were presented first in the museum setting, which did not permit stringent environment controls and where the participants were often surrounded by peers or family members (see [Figures 8D and 8E](#) in [STAR Methods](#)). The subsequent laboratory tests that ensured a carefully controlled environment corroborated and refined these results.

While all four motor tasks exhibited a consistent developmental progression, the results of the four games were largely uncorrelated across individuals (although some significances were seen). An individual that excelled in one task did not necessarily excel in the other tasks. This independence of results indicated that each task measured relatively independent aspects of predictive skills. To our knowledge, the literature to date has treated prediction as a unified concept. More investigation is needed to explore the task-based dissociation of this computational process.

These results add to our understanding of the developmental progression of predictive skills, a core feature in many cognitive processes. Predictive coding has long been discussed as central for perception



Figure 8. Experimental setups for the five games

(A) Schematic of the experimental setup in the laboratory. Vertical paddle movements were mapped onto vertical paddle displacements on the screen. Smaller children stood on an adjustable platform to ensure that the same comfortable arm position corresponded to the desired paddle height on the projector screen. In the museum setup, a computer monitor replaced the screen and back-projection setup.

(B) Button box with one button for the pausing game.

(C) Button box for the choosing game with three buttons, corresponding to the destination locations on the display.

(D) and (E) Scenes from the data collection at the museum.

and cognition as predicted sensory consequences that, when matched with the actual sensory input, yield prediction errors that shape our mental models of the world.^{39–41} A recent computational framework based on predictive coding proposed that predictive learning is fundamental for cognitive development and for acquiring essential behavioral features, such as distinction between self and others, imitation, and goal-directed actions.^{42,43} In the context of “developing robots”, Nagai showed that execution of the action yielded error signals that helped train the initially immature predictor and update the sensorimotor system. However, experimental data in humans that can dissociate and quantify components of predictive skill have so far been quite sparse. We believe that the results reported here help partially address this need.

Given the multitude of studies on Bayesian determinants in human behavior, it needs to be pointed out that our study adopted a different methodological approach to examine prediction. Our study examined single trials where successful interception relies on predicting a dynamic object and synchronizing one’s own hand trajectory with that of an object’s trajectory. In contrast, for example, Mento and Granzio (2020) adopted a Bayesian perspective and studied temporal expectancy in action preparation as a function of preceding statistical properties of the environment.^{44–47} Different probabilistic features created different degrees of bias in a reaction time task. Temporal expectancy and proactive motor control was shaped based on the probabilistic context. A related electroencephalogram study revealed that beta-band desynchronization underlies the motor adjustment based on the probabilistic context.⁴⁶ Similarly, children allocated

attentional resources based on the *a priori* generated temporal expectancy, measured by event-related potentials that were already established in 8- to 12-year-olds.⁴⁷ While interesting and related, this type of prediction is complementary, but nevertheless different from the focus of this study.

Could the results in these perceptually mediated tasks be confounded by the maturation of the visual system that limits the perception of the flying ball? A wide range of studies have examined visual processing of speed in children using different stimuli and procedures. For example, sensitivity to translational motion matures between 8 and 14 years of age.^{48,49} Dynamic visual acuity reaches adult levels at the age of 15 years⁵⁰ and direction discrimination thresholds and global motion perception improve gradually until the age of 14 years.^{51–53} These psychophysical studies also report differences between slow and fast speeds, with younger children having problems perceiving faster speeds.⁵⁴

Even though these data appear to overlap with the developmental curves of our prediction skills, visual maturation is unlikely to account for our findings. The stimuli we used, comprising a single non-noisy high-contrast moving object—the ball—obviated the need for spatial integration across multiple low-contrast entities interspersed with noisy trajectories, as is the case in global motion displays. The observed developmental change^{48,49,51–53} is believed to be related to changes in spatial integration and noise tolerance, neither of which are relevant for our presentations. For example, the developmental progression reported in the study by Schrauf⁵⁰ refers to second-order, non-luminance contrast-defined stimuli which are very different from the punctate stimuli we used. Indeed, as that paper reports, acuity thresholds for high-contrast, non-noisy stimuli are adult-like for even the youngest participants we included in our studies.

To summarize, the predominant paradigm of random dot displays that has been used for studying developmental changes in visual discrimination differs significantly from the task of tracking the trajectory of a single high-contrast discover 400–600 ms. For such punctate stimuli, there is no compelling evidence that low-level visual processes are differential bottlenecks for different ages. Furthermore, if low-level visual processes were the limiting factors driving the age-related changes in our games, we would expect reaction time data to show conformal trends, which was not the case. It is worth noting that in a study by Fischman et al. on ball catching, even very young children (5 years) were sensitive to the perceived aspects of the toss and responded with an appropriate hand orientation, while catch failure was ascribed to coordinative demands by the grasp during the interception.²⁰

Could the development of intuitive physics across the life span explain our results? There is much evidence that humans have an implicit understanding about causality in dynamic events governed by normative physical principles. Research over more than two decades has shown that even young infants have a rich understanding of the physics of the world.^{55–57} For example, studies that separately assessed recognition and prediction of physical events revealed that predictive ability developed later than recognition ability during childhood.^{58–60} The two distinct developmental trajectories imply that understanding current physical events and predicting future physical events have different underlying processes. Furthermore, in the design of our games, we explicitly subjected ball trajectories to gravity values that differed from 9.81 m/s². This not only prevented subjects from relying on intrinsic expectations of the flight trajectories, but it also allowed the ball trajectories to satisfy other constraints. For example, to create identical ball paths with different speeds in the pausing game and suitable time windows to contact the ball in other games, we modified the values of gravity used for the simulations of the ball trajectories (see [STAR Methods](#) for details). Hence, it is unlikely that developmental differences in intuitive understanding of flight dynamics in the natural world accounted entirely for our results.

If not intuitive physics, could successful performance have resulted from learning simple stimulus-response associations? In most games, continuous hand movements had to be temporally and spatially coordinated with an unfolding dynamic event, i.e., the ball's trajectory. As the ball trajectories were varied across trials and launch times were randomized to prevent rote responses, it is not clear how stimulus-response associations might have accounted for improvements with age for the games with parametrically varied stimuli. The possibility of stimulus-response learning may apply to the choosing game, where the point of disappearance of the ball trajectory may become associated with a specific landing box. However, the results exhibited significant location effects, i.e., the lowest target location presented the highest challenge, likely due to the largest spatial difference between trajectory entry and exit height when crossing behind the occlusion. Even if stimulus-response learning affected the

developmental change in the present study, that would not fully explain our results. Previous studies have shown that simple stimulus-response associations rapidly develop over the first three years of life.^{61,62} A study that presented developmental norms for the Wisconsin Card Sorting Test, which involves complex stimulus-response relations, revealed that the performance outcomes at age 10 became indistinguishable from those in healthy young adults.⁶³

Could the results be explained by development of motor coordination abilities involved in our paddle games? A relatively large literature has examined the age-dependent improvements in ball catching, although the metrics were frequently only qualitative and the foci differed widely.^{20,64} Improvements with age have been associated with the visual estimation of the ball's trajectory,^{22,65} response programming,^{66–68} and control and execution of the programmed response.⁶⁹ The functional development of proprioceptive control has been shown to reach adult level by age 8–12 years.²¹ While again these processes overlap with the development of predictive skill, our games explicitly did not require distal motor function by using a paddle rather than a grasp. Furthermore, the pausing and choosing tasks only involved flexions of a single finger without any spatial or force accuracy requirements. Hence, development of fine motor function was not limiting the manifestation of prediction.

These results provide a potential link between earlier reports of the time course of cerebellar maturation and manifest motor behavior. The cerebellum has long been recognized for its prominent role in sensorimotor coordination, but more recently has also been implicated in a wide array of cognitive and emotional functions.^{70–72} Several studies on pediatric cohorts have shown that the gray matter volume in the cerebellum exhibits a U-shape growth trajectory, reaching its peak size around age 11.^{73,74} The inferior posterior lobe of the cerebellum that is associated with cognitive functions reaches its peak volume in parallel with volumetric changes in the gray matter of the frontal lobe. While speculative, this result shows notable consonance with the developmental trends in our behavioral data.

This work points to several additional promising avenues for further research. First, our cross-sectional dataset investigating predictive abilities in healthy individuals across the life span may serve as normative reference to understand potential deviations from the typical developmental trajectory. For example, a growing body of literature suggests that autistic individuals have divergent predictive abilities that could underlie a wide array of symptoms crossing the cognitive, social, sensory, and motor domain.^{36–38,42,75} In fact, this study was originally motivated by a new theoretical perspective on understanding the mechanisms underlying the wide spectrum of differences in ASD: the hypothesis that the observable features of autism arise due to predictive impairments.^{36–38,75} The current research is a step toward testing whether predictive ability can be measured with our custom-developed games in a typically developing population. Future investigations will explore whether autistic individuals exhibit systematic performance differences in the suite of predictive tasks presented here. Given the acknowledged role of the cerebellum in predictive skills, it is notable that several past studies of brain structure have reported differences in cerebellar volume in autistic individuals relative to their non-autistic counterparts.^{76,77}

Furthermore, given the centrality of prediction in all motor, cognitive, and social skills, one eminent question is whether prediction can be improved through practice. Following Nagai's computational framework, postulating that sensorimotor prediction is foundational for cognitive and social development, sensorimotor training might be a context which could boost predictive skills in other domains.⁴² This question cannot be addressed by our current data as our sessions were relatively short (6–9 min per task in the laboratory, and even shorter in the museum), leaving no opportunity to examine practice effects. Examining learning and skill retention could have translational applications for clinical populations where sensorimotor prediction is affected.

A final question is to what degree sex differences exist, as has been amply documented in the literature on ball skills. In throwing and catching, there appears to be about a one-year performance differential between the sexes, with girls tending to lag boys, although practice may play a big role.^{20,78,79} While this advance seems to contradict the advance in cerebellar maturation, the question remains whether analogous sex- or gender-based differences exist in predictive abilities.

Limitations of the study

While encouraged by the cross-task and cross-setting consistencies in our results, we acknowledge that the cross-sectional design has some limitations.⁸⁰ The discretization into age groups that were necessary for

comparing cross-sectional groups is a case in point. Future studies could increase the age resolution to scrutinize the transition to adult-like performance in a more continuous manner.

Furthermore, the age groups in the laboratory and museum setting have different distributions. This is the result of restrictions imposed by this study in the real world. Data were collected on numerous weekends over 8 months and the museum did not allow any selective participant recruitment to conform to any specific demographic distributions. Hence, the main portion of the experiment was repeated in our laboratory under maximally controlled conditions. Note that having the same participants perform both experiments was also infeasible due to museum policy forbidding separately recontacting families.

Notwithstanding these challenges, it is worth noting that the data from the two settings are broadly consistent and complementary. The museum setting allowed only limited control over extraneous distractions. However, verifying the consistency of the museum-based data with those from the better controlled laboratory setting allowed us to mitigate the concern that our inferences may be biased by the methodological shortcomings of the former. As a more general guiding principle, we believe that results that are as subtle as to only reveal themselves under maximal experimental control may fall short on the criterion of ecological validity.

While this study examined movement games, the experimental and theoretical focus was on the development of a computational ability, prediction, not on motor coordination per se. The different interception tasks were carefully designed to carve out prediction as the common thread across the four games. This is evident from the choice of reaction time as control task, using very similar motor coordination, but with markedly different timing demands. And yet, even if prediction is core to many other domain tasks, ranging from perception to cognition, it is premature to extrapolate and speculate about prediction in cognition in general. More work is needed to establish how shared or disjoint predictive processes are across different task domains.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHODS DETAILS
 - Experimental equipment
- EXPERIMENTAL PROCEDURE
 - Experimental task details
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106038>.

ACKNOWLEDGMENTS

We first want to thank all participants and their families who participated in the laboratory and museum studies for their time and cooperation. We would like to express our gratitude to Dr. Robert Joseph for advising on inclusion/exclusion criteria and characterization of participants, and Kelly Chan, Leila Denna, and Amanda J. Haskins for their assistance with participant recruitment, characterization, and data collection. We also want to thank the many student volunteers who readily and enthusiastically spent their weekends at the museum to recruit participants for our study: undergraduate students: Hannah Tam, Kaleb Noruzi, Ellesse Cooper, Anne Sigl, Jeffrey Zhu, Abigail Cahill, Sasha Dubinsky, Dena Crozier; graduate students: Rashida Nayeem and Zhaoran Zhang. We also want to thank Michael Graham, Salah Bazzi and Jim Papadopoulos for designing and building the paddle device customized for the museum data collection. The Museum of Science granted support to the lead undergraduate student, Hannah Tam. The study was funded by the Simons Foundation Autism Research Initiative, SFARI-192901 (D.S. and P.S.), the Simons

Center for the Social Brain at MIT (P.S. and M.K.) and the National Institutes of Health, R21-HD089731 (D.S. and P.S.) and R37-HD087089 (D.S.).

AUTHOR CONTRIBUTIONS

Conceptualization: S.W.P., A.C., P.S., and D.S.; Methodology: S.W.P., A.C., P.S., D.S., and M.K.; Investigation: S.W.P., A.C., and D.C.; Visualization: S.W.P., A.C., and P.S.; Statistical analysis: M.R. and S.W.P.; Supervision: P.S. and D.S.; Writing—original draft: S.W.P., A.C., P.S., and D.S.; Writing—review & editing: S.W.P., A.C., D.C., M.R., M.K., P.S., and D.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: May 2, 2022

Revised: December 14, 2022

Accepted: January 17, 2023

Published: January 25, 2023

REFERENCES

- Wolpert, D.M., Diedrichsen, J., and Flanagan, J.R. (2011). Principles of sensorimotor learning. *Nat. Rev. Neurosci.* 12, 739–751.
- Wolpert, D.M., and Flanagan, J.R. (2001). Motor prediction. *Curr. Biol.* 11, R729–R732.
- Zago, M., McIntyre, J., Senot, P., and Lacquaniti, F. (2009). Visuo-motor coordination and internal models for object interception. *Exp. Brain Res.* 192, 571–604.
- Aruin, A.S., Forrest, W.R., and Latash, M.L. (1998). Anticipatory postural adjustments in conditions of postural instability. *Electroencephalogr. Clin. Neurophysiol.* 109, 350–359.
- Blakemore, S.-J., Wolpert, D.M., and Frith, C.D. (1998). Central cancellation of self-produced tickle sensation. *Nat. Neurosci.* 1, 635–640.
- Maurice, P., Hogan, N., and Sternad, D. (2018). Predictability, force, and (anti) resonance in complex object control. *J. Neurophysiol.* 120, 765–780.
- Nasserolelami, B., Hasson, C.J., and Sternad, D. (2014). Rhythmic manipulation of objects with complex dynamics: predictability over chaos. *PLoS Comput. Biol.* 10, e1003900.
- Nayeem, R., Bazzi, S., Sadeghi, M., Hogan, N., and Sternad, D. (2021). Preparing to move: setting initial conditions to simplify interactions with complex objects. *PLoS Comput. Biol.* 17, e1009597.
- Parr, T., Pezzulo, G., and Friston, K.J. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior* (MIT Press).
- Ciria, A., Schillaci, G., Pezzulo, G., Hafner, V.V., and Lara, B. (2021). Predictive processing in cognitive robotics: a review. *Neural Comput.* 33, 1402–1432.
- Spratling, M.W. (2017). A review of predictive coding algorithms. *Brain Cogn.* 112, 92–97.
- Huang, Y., and Rao, R.P.N. (2011). Predictive coding. *Wiley Interdiscip. Rev. Cogn. Sci.* 2, 580–593.
- Körding, K.P., and Wolpert, D.M. (2004). Bayesian integration in sensorimotor learning. *Nature* 427, 244–247.
- Chater, N., Oaksford, M., Hahn, U., and Heit, E. (2010). Bayesian models of cognition. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 811–823.
- Aitchison, L., and Lengyel, M. (2017). With or without you: predictive coding and Bayesian inference in the brain. *Curr. Opin. Neurobiol.* 46, 219–227.
- Clark, J.E., and Metcalfe, J.S. (2002). In *Motor Development: Research & Reviews*, J.E. Clark and J.H. Humphrey, eds. (NASPE Publications), pp. 163–180.
- Haywood, K.M., and Getchell, N. (2020). *Life Span Motor Development* (Human Kinetics).
- McKay, M.J., Baldwin, J.N., Ferreira, P., Simic, M., Vanicek, N., and Burns, J.; 1000 Norms Project Consortium (2017). Reference values for developing responsive functional outcome measures across the lifespan. *Neurology* 88, 1512–1519.
- van Beurden, E., Zask, A., Barnett, L.M., and Dietrich, U.C. (2002). Fundamental movement skills—how do primary school children perform? The ‘Move it Groove it’ program in rural Australia. *J. Sci. Med. Sport* 5, 244–252.
- Fischman, M.G., Moore, J.B., and Steele, K.H. (1992). Children’s one-hand catching as a function of age, gender, and ball location. *Res. Q. Exerc. Sport* 63, 349–355.
- Gabbard, C. (2021). *Lifelong Motor Development* (Lippincott Williams & Wilkins).
- Williams, J.G. (1992). Catching action: visuomotor adaptations in children. *Percept. Mot. Skills* 75, 211–219.
- von Hofsten, C. (1983). Catching skills in infancy. *J. Exp. Psychol. Hum. Percept. Perform.* 9, 75–85.
- Contreras-Vidal, J.L. (2006). Development of forward models for hand localization and movement control in 6- to 10-year-old children. *Hum. Mov. Sci.* 25, 634–645.
- Contreras-Vidal, J.L., Bo, J., Boudreau, J.P., and Clark, J.E. (2005). Development of visuomotor representations for hand movement in young children. *Exp. Brain Res.* 162, 155–164.
- Kagerer, F.A., Contreras-Vidal, J.L., Bo, J., and Clark, J.E. (2006). Abrupt, but not gradual visuomotor distortion facilitates adaptation in children with developmental coordination disorder. *Hum. Mov. Sci.* 25, 622–633.
- Kagerer, F.A., and Clark, J.E. (2014). Development of interactions between sensorimotor representations in school-aged children. *Hum. Mov. Sci.* 34, 164–177.
- Janacsek, K., Fiser, J., and Nemeth, D. (2012). The best time to acquire new skills: age-related differences in implicit sequence learning across the human lifespan. *Dev. Sci.* 15, 496–505.
- Du, Y., Valentini, N.C., Kim, M.J., Whithall, J., and Clark, J.E. (2017). Children and adults both learn motor sequences quickly, but do so differently. *Front. Psychol.* 8, 158.

30. Ozonoff, S., and Jensen, J. (1999). Brief report: specific executive function profiles in three neurodevelopmental disorders. *J. Autism Dev. Disord.* 29, 171–177.
31. Palisano, R.J., Hanna, S.E., Rosenbaum, P.L., Russell, D.J., Walter, S.D., Wood, E.P., Raina, P.S., and Galuppi, B.E. (2000). Validation of a model of gross motor function for children with cerebral palsy. *Phys. Ther.* 80, 974–985.
32. Skinner, R.A., and Piek, J.P. (2001). Psychosocial implications of poor motor coordination in children and adolescents. *Hum. Mov. Sci.* 20, 73–94.
33. Fournier, K.A., Hass, C.J., Naik, S.K., Lodha, N., and Cauraugh, J.H. (2010). Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *J. Autism Dev. Disord.* 40, 1227–1240.
34. Ghaziuddin, M., and Butler, E. (1998). Clumsiness in autism and Asperger syndrome: a further report. *J. Intellect. Disabil. Res.* 42, 43–48.
35. Gowen, E., and Hamilton, A. (2013). Motor abilities in autism: a review using a computational context. *J. Autism Dev. Disord.* 43, 323–344.
36. Cannon, J., O'Brien, A.M., Bungert, L., and Sinha, P. (2021). Prediction in autism spectrum disorder: a systematic review of empirical evidence. *Autism Res.* 14, 604–630.
37. Sinha, P., Kjelgaard, M.M., Gandhi, T.K., Tsourides, K., Cardinaux, A.L., Pantazis, D., Diamond, S.P., and Held, R.M. (2014). Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. USA.* 111, 15220–15225.
38. Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de Wit, L., and Wagemans, J. (2014). Precise minds in uncertain worlds: predictive coding in autism. *Psychol. Rev.* 121, 649–675.
39. Adams, R.A., Shipp, S., and Friston, K.J. (2013). Predictions not commands: active inference in the motor system. *Brain Struct. Funct.* 218, 611–643.
40. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204.
41. Koster-Hale, J., and Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron* 79, 836–848.
42. Nagai, Y. (2019). Predictive learning: its key role in early cognitive development. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374, 20180030.
43. Philippssen, A., and Nagai, Y. (2020). A predictive coding account for cognition in human children and chimpanzees: a case study of drawing. *IEEE Trans. Cogn. Dev. Syst.* 14, 1306–1319.
44. Duma, G.M., Granzol, U., and Mento, G. (2020). Should I stay or should I go? How local-global implicit temporal expectancy shapes proactive motor control: an hEEG study. *Neuroimage* 220, 117071.
45. Mento, G., and Granzol, U. (2020). The developing predictive brain: how implicit temporal expectancy induced by local and global prediction shapes action preparation across development. *Dev. Sci.* 23, e12954.
46. Mento, G., Scerif, G., Granzol, U., Franzoi, M., and Lanfranchi, S. (2020). The effect of probabilistic context on implicit temporal expectations in down syndrome. *Front. Psychol.* 11, 369.
47. Mento, G., and Vallesi, A. (2016). Spatiotemporally dissociable neural signatures for generating and updating expectation over time in children: a high density-ERP study. *Dev. Cogn. Neurosci.* 19, 98–106.
48. Bogfjellmo, L.-G., Bex, P.J., and Falkenberg, H.K. (2014). The development of global motion discrimination in school aged children. *J. Vis.* 14, 19.
49. Narasimhan, S., and Giaschi, D. (2012). The effect of dot speed and density on the development of global motion perception. *Vision Res.* 62, 102–107.
50. Schrauf, M., Wist, E.R., and Ehrenstein, W.H. (1999). Development of dynamic vision based on motion contrast. *Exp. Brain Res.* 124, 469–473.
51. Ellemberg, D., Lewis, T.L., Liu, C.H., and Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Res.* 39, 2325–2333.
52. Gunn, A., Cory, E., Atkinson, J., Braddick, O., Wattam-Bell, J., Guzzetta, A., and Cioni, G. (2002). Dorsal and ventral stream sensitivity in normal development and hemiplegia. *Neuroreport* 13, 843–847.
53. Parrish, E.E., Giaschi, D.E., Boden, C., and Dougherty, R. (2005). The maturation of form and motion perception in school age children. *Vision Res.* 45, 827–837.
54. Manning, C., Aagten-Murphy, D., and Pellicano, E. (2012). The development of speed discrimination abilities. *Vision Res.* 70, 27–33.
55. Spelke, E.S., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). Origins of knowledge. *Psychol. Rev.* 99, 605–632.
56. Michotte, A. (1963). *The Perception of Causality* (Methuen & Co).
57. Kubricht, J.R., Holyoak, K.J., and Lu, H. (2017). Intuitive physics: current research and controversies. *Trends Cogn. Sci.* 21, 749–759.
58. Kim, I., and Spelke, E.S. (1999). Perception and understanding of effects of gravity and inertia on object motion. *Dev. Sci.* 2, 339–362.
59. Howe, C., Taylor Tavares, J., and Devine, A. (2014). Children's conceptions of physical events: Explicit and tacit understanding of horizontal motion. *Br. J. Dev. Psychol.* 32, 141–162.
60. Hast, M., and Howe, C. (2017). Changing predictions, stable recognition: children's representations of downward incline motion. *Br. J. Dev. Psychol.* 35, 516–530.
61. Zelazo, P.D., and Reznick, J.S. (1991). Age-related asynchrony of knowledge and action. *Child Dev.* 62, 719–735.
62. Bunge, S.A., and Zelazo, P.D. (2006). A brain-based account of the development of rule use in childhood. *Curr. Dir. Psychol. Sci.* 15, 118–121.
63. Chelune, G.J., and Baer, R.A. (1986). Developmental norms for the Wisconsin card sorting test. *J. Clin. Exp. Neuropsychol.* 8, 219–228.
64. Ripoll, H., Keller, J., and Olivier, I. (1994). Le développement du comportement moteur de l'enfant: l'exemple des saisies et des interceptions de balle. *enfan.* 47, 265–284.
65. Abernethy, B. (1988). The effects of age and expertise upon perceptual skill development in a racquet sport. *Res. Q. Exerc. Sport* 59, 210–221.
66. Clark, J.E. (1982). Developmental differences in response processing. *J. Mot. Behav.* 14, 247–254.
67. Olivier, I., Ripoll, H., and Audiffren, M. (1996). The development of initiating and execution of ball interception in children/Ontogenèse des mécanismes de traitement de l'information impliqués dans le déclenchement et l'exécution d'un mouvement d'interception. *Sci. Mot.: revue scientifique de l'ACAPS/ACAPS* 27, 11–16.
68. Reilly, M.A., and Spirduso, W.W. (1991). Age-related differences in response programming. *Res. Q. Exerc. Sport* 62, 178–186.
69. Keller, J. (1992). Activité physique et sportive et motricité de l'enfant [Physical and Sporting Activity and Motor Behavior of Children (Vigot)].
70. Middleton, F.A., and Strick, P.L. (1997). Cerebellar output channels. *Int. Rev. Neurobiol.* 41, 61–82.
71. Riva, D., and Giorgi, C. (2000). The cerebellum contributes to higher functions during development: evidence from a series of children surgically treated for posterior fossa tumours. *Brain* 123, 1051–1061.
72. Schmahmann, J.D. (2004). Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. *J. Neuropsychiatry Clin. Neurosci.* 16, 367–378.
73. Brain Development Cooperative Group (2012). Total and regional brain volumes in a population-based normative sample from 4 to 18 years: the NIH MRI Study of Normal Brain Development. *Cerebr. Cortex* 22, 1–12.
74. Tiemeier, H., Lenroot, R.K., Greenstein, D.K., Tran, L., Pierson, R., and Giedd, J.N. (2010). Cerebellum development during childhood and adolescence: a longitudinal morphometric MRI study. *Neuroimage* 49, 63–70.

75. Lawson, R.P., Rees, G., and Friston, K.J. (2014). An aberrant precision account of autism. *Front. Hum. Neurosci.* 8, 302.
76. Stanfield, A.C., McIntosh, A.M., Spencer, M.D., Philip, R., Gaur, S., and Lawrie, S.M. (2008). Towards a neuroanatomy of autism: a systematic review and meta-analysis of structural magnetic resonance imaging studies. *Eur. Psychiatry* 23, 289–299.
77. Valera, E.M., Faraone, S.V., Murray, K.E., and Seidman, L.J. (2007). Meta-analysis of structural imaging findings in attention-deficit/hyperactivity disorder. *Biol. Psychiatry* 61, 1361–1369.
78. McConnell, A., and Wade, G. (1990). Effects of lateral ball location, grade, and sex on catching. *Percept. Mot. Skills* 70, 59–66.
79. Crozier, D., Zhang, Z., Park, S.-W., and Sternad, D. (2019). Gender differences in throwing revisited: sensorimotor coordination in a virtual ball aiming task. *Front. Hum. Neurosci.* 13, 231.
80. Adolph, K.E., Berger, S.E., and Leo, A.J. (2011). Developmental continuity? Crawling, cruising, and walking. *Dev. Sci.* 14, 306–318.
81. Kleiner, M., Brainard, D., and Pelli, D. (2007). What is new in Psychophysics Toolbox. *Perception* 36, 416.
82. Huber, M.E., and Sternad, D. (2015). Implicit guidance to stable performance in a rhythmic perceptual-motor skill. *Exp. Brain Res.* 233, 1783–1799.
83. Wei, K., Dijkstra, T.M.H., and Sternad, D. (2007). Passive stability and active control in a rhythmic task. *J. Neurophysiol.* 98, 2633–2646.
84. Baayen, R.H., Davidson, D.J., and Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412.
85. Barr, D.J. (2013). Random Effects Structure for Testing Interactions in Linear Mixed-Effects Models (Frontiers Media SA).
86. Cesqui, B., Russo, M., Lacquaniti, F., and d'Avella, A. (2016). Grasping in one-handed catching in relation to performance. *PLoS One* 11, e0158606.
87. Russo, M., Cesqui, B., La Scaleia, B., Ceccarelli, F., Maselli, A., Moscatelli, A., Zago, M., Lacquaniti, F., and d'Avella, A. (2017). Intercepting virtual balls approaching under different gravity conditions: evidence for spatial prediction. *J. Neurophysiol.* 118, 2421–2434.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
De-identified performance data from human subjects	This paper	https://github.com/swpark7948/PredictionDevelopment
Analysis codes	This paper	https://github.com/swpark7948/PredictionDevelopment
Software and algorithms		
MATLAB 2016b	Mathworks	https://www.mathworks.com/
Psychophysics Toolbox-3	Kleiner, et al. (2007) ⁸¹	http://psychtoolbox.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Se-Woong Park (sewoong.park@utsa.edu).

Materials availability

Visual stimuli are shown in the [supplemental information](#) file.

Data and code availability

De-identified performance data from human subjects and the analysis codes are available from <https://github.com/swpark7948/PredictionDevelopment>

This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A first series of data collection took place in the Living Laboratory at the Museum of Science in Boston as part of a science outreach activity. Over the course of 8 months, 442 museum visitors (233 males and 209 females) participated in the study (Figures 8D and 8E). Their age range was 5 to 92 years, with a median age of 18.7 years. As this was an outreach activity, participants were volunteers and the age range and distribution could not be influenced. Each participant or their legal guardian filled out a short questionnaire about age, sex, handedness, and history of movement, neurological, developmental, or psychiatric disorders. 51 out of the 442 participants who self-reported a movement or psychiatric disorder in the questionnaire were excluded from analyses, leaving a total of 391 participants. After an additional exclusion of 21 participants who provided irrelevant or insufficient questionnaire information, a total of 370 participants (199 males and 171 females) could be used for data analysis. Due to the spontaneous nature of recruitment and limited participation time per person, not all participants completed all five games. The number of participants per game thus varied across the five games (n=288, 286, 322, 317, 269, in order of the games presented in the results above).

For the laboratory experiment, a cohort of 30 typically developing children and 14 healthy adults was recruited. All children were between the ages of 7 and 12 years (17 males, 13 females, 9.5 ± 1.7 years); the adults (8 males, 6 females) were college students between the ages of 18 and 21 years (19.8 ± 1.1 years). Individuals who reported a history of neurological, movement, neurodevelopmental, or psychiatric diagnoses in the initial phone screening were excluded.

All participants or their legal guardians (if under 18 years) signed the consent form (and assent for minors) before starting the experiments. All procedures were approved by the Institutional Review Boards of Northeastern University, the Massachusetts Institute of Technology, and the Museum of Science, Boston.

METHODS DETAILS

Experimental equipment

Experimental equipment for recording paddle movements

The catching, bouncing, and reaction time games involved a custom-designed paddle interfaced with the virtual environment (screen). Participants grasped the handle of a real ping pong paddle with their dominant hand (Figures 1F, 8A, 8D and 8E). To play the game, they moved the paddle upwards to catch or bounce a ball. The paddle movements were measured by a metal rod attached to the bottom face of the paddle that slid through a loop and contacted the gear of an optical encoder (Bourns Inc., Riverside, CA). The encoder signal was collected by a data acquisition board at a sampling rate of 1kHz with a spatial resolution of 0.27mm (NI-USB 6343, National Instruments, Austin, TX). Two hinge joints in the rod allowed participants to move the racket freely in 3D with minimal friction limiting the movements, although only the vertical component of the paddle motion generated the movement of the virtual racket. As the display only showed vertical paddle movements, participants tended to and were instructed to stay close to vertical movements. The visuo-motor delay between the real and visually displayed paddle movements was 22ms. The interface was programmed in MATLAB Psychophysics Toolbox v.3 (The Mathworks, Natick, MA). This virtual set-up has been validated in previous experiments with different graphic interfaces.^{82,83}

For data collection at the museum, the device had to be robust and easy to assemble as children tended to congregate around the device and unintended contacts with the device occurred (Figure 8E). Hence, a customized version of the paddle device was built for museum use with additional plastic casing around the encoder. Instead of a PC computer and a back projection screen, a laptop computer connected to an LCD monitor (75cm width and 45cm height) and a USB-type data acquisition board (NI-USB 6343) were used at the museum.

Experimental equipment for recording button presses

Participants performed two tasks interacting with a virtual display via pressing a button to indicate time and location of a ball. Two button boxes were made of mint tins (6.0cm x 9.5cm x 3.5cm). The first box contained one large red button (3.0cm diameter) and was used for the pausing game to test temporal prediction (Figure 8B). The second box contained three small buttons (1.6cm diameter) in red, green, and blue color that were used for the choosing game to test spatial prediction (Figure 8C). The buttons in both boxes were wired to connect with the same data acquisition board (NI-6289) at a sampling rate of 1kHz to afford interaction with the virtual events on the screen. A PC computer (Dell Precision T1500) controlled this interaction and the interface was programmed in the MATLAB Psychophysics Toolbox v.3 and enabled real-time feedback with the virtual display.⁸¹ The same equipment was used for data collection at the museum, except that the experiments were run by a laptop computer connected to an LCD monitor (75cm width and 45cm height) and a USB-type data acquisition board (NI-USB 6343).

EXPERIMENTAL PROCEDURE

For the data collection at the museum, two experimental stations were set up in a designated booth at the Living Laboratory. Our outreach research event took place every Saturday for 3 hours over 8 months, involving a team of 12 student volunteers from Northeastern University. For each shift, 3 out of the 12 students were present: one person recruited participants from the adjacent museum exhibits and provided information to museumgoers about the scientific questions as an educational activity; the second person explained the study and obtained informed consent, handed out and collected the questionnaire; the third person oversaw the data collection. Volunteers were trained to follow the same protocol with the same verbal instructions with each participant. The games with the paddle and the button presses ran simultaneously at two adjacent stations (Figures 8D and 8E). Height-adjustable steps were used for smaller children to ensure that the participants played with a comfortable arm configuration (Figure 8E). The LCD monitor displayed the games at approximately 75cm distance. As the pictures convey, the museum was far from a controlled environment as peers and family frequently watched and waited for their turn. However, this prior observation helped to reduce the instruction time, as participation time at the museum was limited to 15 minutes per person.

For data collection in the laboratory, participants sat or stood in front of a large back projection screen (3.0 × 2.5m) which displayed the games at approximately 1.50m distance (Figures 1G and 8A). Participants stood on an height-adjustable pedestal to ensure maximum ease in manipulating the paddle (Figure 8A).

For the two button tasks, participants sat in a chair with the button boxes on a table. Each game was explained by a standardized procedure using slides with a detailed pre-recorded instructional audio narrative together with displays of the games. Subsequently, the experimenter demonstrated five trials of the task, followed by the participant practicing five trials before commencing data collection for each game.

Experimental task details

The following section includes images of the actual task stimuli (rather than schematics) as well as detailed descriptions of how the stimuli were generated and the metrics defined (velocity, acceleration, duration, spatial trajectory, and trial-to-trial variability). This information aims to promote in-depth interpretation of the results and to enable replication by other groups. We specify the dimensions of visual stimuli based on the experimental setup (back projection screen) in the laboratory. Since we used computer monitors in the museum, a scale factor of 0.25 is applied for the visual stimuli in the museum, i.e., the display in the laboratory is 4 times larger than that in the museum.

Pilot testing

All experimental tasks were designed and tuned in extensive pilot testing in the laboratory with typically developing children and young adults. This testing aimed to ensure that the challenges for prediction and motor actions were manageable, while also engaging, especially for children aged 7–12 years.

Task theme and instructions

The unifying theme among the five games was mice seeking cheese; the instruction and feedback slides between the games included images of the stimuli as participants saw them during the tasks. Before completing the tasks, participants at the laboratory received general instructions about the setup and what they would be asked to do. All participants were shown a visual schedule with a picture and title of each task and offered a sticker after completing each experiment. To ensure comprehension for all age groups, instructions for all games followed a consistent format: 1) labeling all components of the game stimuli, 2) stating the goal of the game, 3) depicting static images of each step in a trial (cue sound, launch sound, stimulus appearing, response required, feedback provided) along with accompanying narrative and game sounds, 4) observing the researcher demonstrating five trials, followed by the participant practicing five trials. Then, participants performed the full block of trials independently. The main instruction was repeated several times during this process for each game, to ensure comprehension particularly with younger children. The same instructions were given to participants at the museum, except that the researcher verbally delivered the instructions without slides and demonstrated three trials prior to three practice trials. A visual schedule was not provided at the museum.

Catching

This first game tested participants' ability to predict and catch a virtual ball with a virtual basket, i.e., by moving the hand-held paddle. Each trial began with the participant placing the paddle at its lowest position within a blue box displayed on the right side of the screen. The yellow ball ('cheese') was launched from another blue box on the left side and flew with a ballistic trajectory to the right. A continuous rattling sound provided an attentional cue before the ball was launched, accompanied by a popping sound at flight onset. The duration of the rattling sound was randomized between 1.5 and 2.5s to render the exact onset of the launch unpredictable. This was intended to prevent stereotyped movements from trial to trial and promote engagement. The ball trajectory followed the equation for ballistic flight in two dimensions, $x = v_x t$, $y = v_y t - \frac{1}{2} a t^2$. To generate two different ball trajectories with the same flight duration, the vertical velocity and acceleration took one of two values: in 50% of trials the vertical velocity was 131cm/s and the acceleration was 330cm/s²; in another 50% of trials the velocity was 106cm/s and acceleration was 270cm/s². For both parameterizations the horizontal velocity component was 131cm/s. The two ball trajectories were randomized to further avoid stereotyped actions and to require real-time prediction. For both museum and laboratory data collection, the ratio of the displacement of the basket on the screen to the real paddle displacement was 1:1.5 (the displacement on the screen was larger).

Participants tried to catch the ball by moving their hand-held paddle/basket upward following ball launch. The width of the basket was 20cm. The goal was to catch the ball as closely as possible in the center of the basket, indicated with a red dot (Figure S1). In addition, the trajectory could be occluded by one, two, or three trees to increase the challenge. Participants received visual and auditory feedback after each trial.

Three distinct audio cues signaled if the ball was missed, caught in the basket, or caught in the center of the basket; in the latter case, three red stars appeared to provide additional reward and motivation. When the paddle was returned to its lowest position inside the blue box, the ball appeared in the left box and the next trial began.

The lateral spatial distance of the ball contact to the basket center quantified the error. If the ball was not caught, the distance between the basket center and the location where the ball passed the top of the basket was calculated. At the museum, due to the time constraints, participants only played 5 trials without trees and subsequently 15 trials with trees; the number of trees was randomized with an equal number of trials per tree number. In the laboratory, the game started with 30 trials without trees, followed by 2 blocks of 30 trials with trees; the number of trees was randomized within each block. For both museum and laboratory data, trials without occlusion were not included in the data analysis.

Bouncing

This game tested the predictive ability of participants to intercept a ball with a paddle and then bounce it into the crocodile's mouth (Figure S2). The red tongue tip in the center of the mouth was the target. This task not only required the timed contact of the launched ball, but also necessitated a specific velocity of the paddle at contact to generate the desired ball trajectory after contact. As before, the yellow ball was launched from the blue box on the left side of the screen. Two parameterized ball trajectories, identical to those in the catching game above, were presented in random order. Participants began with the paddle, a horizontal bar, at its lowest position inside the blue box and, upon appearance of the ball, attempted to bounce the ball with a vertical movement of the paddle. After each bounce, three distinct visual and auditory cues signaled whether the ball missed the mouth, landed inside the mouth, or hit the tip of the crocodile's tongue; the latter was additionally rewarded with a slurping sound. The accuracy was defined along the imaginary line passing through the upper and lower jaws of the crocodile. The absolute distance of where the ball crossed this line and the tip of the tongue served as the error measure. If the ball missed the mouth, the same line was extended and the distance between contact and tongue tip was calculated. At the museum, participants practiced for 3 trials and played 15 trials. In the laboratory, participants practiced for 5 trials and then completed two blocks of 30 trials, with 15 trials per ball trajectory type.

Pausing

This game tested temporal prediction using the button box with a single large button (Figure 8B). To start the trial, the same rattling sound provided a cue prior to ball launch, accompanied by a popping sound when the ball left the box. The duration of the rattling sound was randomized between 1.5 and 2.5s to avoid predictable timing of the launch and necessitate online prediction of the trajectory. The ball was launched from a blue box on the left side of the screen and flew to the right following a ballistic trajectory (Figure S3A). The horizontal distance between the launch box and the center of the window was 58cm; the size of the ball on the screen was 3.5cm. The participants could see the flying ball until it disappeared inside the gray house (width = 25cm). They were instructed to press the red button when they anticipated the ball to reappear in the window of the house ("where a mouse lives"). Crosshairs (intersecting window muntins) identified the center of the window (width=10cm). After pressing the button, the ball stopped and its location in the window or behind the house was revealed to the participant to provide feedback about accuracy. Temporal error was defined as the difference between the time of the actual stop and the time when the ball would have been exactly in the crosshairs of the window.

The same flight equation was used to calculate three ball trajectories with three different horizontal velocities v_x : 105, 117, 131cm/s, determining the durations of the ball flight to the target to be 552, 497 and 442ms, respectively. Three corresponding vertical velocities and accelerations were chosen such that the ball always followed the same ballistic path despite the different flight durations; v_y : 80, 89, 100cm/s and a_y : 160, 198, 250cm/s². This parameterization eliminated any spatial cues to predict the ball's trajectory. For further feedback three different sounds were played when the ball stopped outside the window, appearing in the window, or exactly in the crosshairs of the window. In the data collection in both the museum and the laboratory, participants performed a block of 30 trials with 10 trials for each ball speed presented in a randomized fashion.

Choosing

This game tested ability to predict where the ball would land without requiring a temporally accurate response. Responses were obtained using the box with three buttons to indicate the landing position of

the ball ('cheese', Figure S3B). The ball was launched from the same blue box as before and followed a similar flight trajectory as in the previous game. However, now three targets at the rightmost side of the house were positioned in a vertical stack ("three mice were waiting for cheese"). To vary the ball's path, only the vertical velocity component varied across trials such that the ball landed at one of the three targets (v_x : 117cm/s, v_y : 80, 89, 100cm/s and a_y : 198cm/s²). The horizontal distance between the launcher and the targets was 83cm. The stacked targets in red, green, and blue corresponded spatially to the three colored buttons. After the ball was launched, its trajectory disappeared for 385ms behind the house and did not reappear until the participant responded. Participants were instructed to indicate at which target the ball landed by pressing the button with the same color (which mouse had "eaten" the cheese). For performance feedback, the target that the participant selected became brighter and simultaneously the 'cheese' also appeared over the target where it actually landed. A correct response was rewarded by a sound. Note that the three ball trajectories differed in their amount of vertical displacement before and after occlusion. The blue trajectory with the largest displacement proved to be the most challenging trajectory.

The task did not require temporally precise or rapid responses. However, if a key was not pressed within 5s, the participant received feedback depicting the correct response. A new trial started automatically within 1s following the feedback regardless of participant response (or lack thereof). Performance was measured only categorically: the success rate was calculated by dividing the number of correct trials by the total number of trials. In both museum and laboratory, participants performed a block of 30 trials with 10 trials for each target location presented in a randomized fashion.

Reacting

This game served as a control task as it tested participants' reaction time or information processing speed. The game was designed to look very similar to the other games to render the reaction time measures comparable. Therefore, participants used the same paddle (represented as a basket), starting each trial with the basket in the same blue box (Figure S4). They were instructed to "catch a mouse" as quickly as possible when it appeared on the screen to "prevent it from eating a wheel of cheese." Each trial began with a continuous drum roll attentional cue sound lasting between 1 to 2s, which signaled to wait. After this random cue interval, a mouse appeared 27cm above a basket (controlled by the paddle). Upon appearance of the mouse, the participants had to lift the paddle as fast as possible and catch it. Reaction time was the time between the appearance of the mouse and the initiation of the paddle movement. At the same time as the mouse appeared, the cheese wheel depicted elapsing time by progressively disappearing wedge by wedge (Figure S4); it completely disappeared after 425ms. The size of the remaining wheel indicated the reaction time: if the mouse was caught faster, more cheese was "saved."

To eliminate trials with a false start or delayed response due to lack of attention, only the trials with the reaction time between 150 and 500ms were analyzed. This range was determined in pilot tests with young adults and children. At the museum data collection, 10 trials were collected; in the laboratory, 20 trials were collected.

QUANTIFICATION AND STATISTICAL ANALYSIS

To assure the quality of the collected data, video recordings were obtained in the laboratory and screened after the data collection to eliminate trials when the participant was looking away, talking, or otherwise visibly not attending to the screen. At the museum, video recording was not allowed and thus post hoc screening was not possible. For both museum and laboratory data, trials with scores outside three standard deviations within each participant were excluded. The overall percentage of the eliminated trials was less than 1%.

For all performance metrics, the full set of museum data was examined first to learn to what extent performance and predictive ability changed over the entire age range from 5 to 92 years of age, even though very few participants over 60 years volunteered. To better visualize the dependent variable of each game with respect to age, individual participant means were fitted with an exponential function, $y = ae^{-bx} + c$, where y was the dependent variable and x was the independent variable, age.

The next step aimed to identify at what age children's performance approached that of adults. The children were parsed into two-year age groups to afford comparison with the adult group. At the museum, the age range was broad and included children as young as 5 years of age. Hence, the museum data had a total of

six child groups: 5–6, 7–8, 9–10, 11–12, 13–14 and 15–16 years of age. This grouping was chosen as it was a compromise between getting a reasonable age resolution and assuring a relatively similar number of samples in each group. The adult group for the museum data comprised individuals 18–21 years of age. The number of participants per group varied across groups and also across tasks as not all participants performed all five games. For the laboratory experiment, the children cohort was split into three child groups: 7–8 ($n = 12$), 9–10 ($n = 12$), and 11–12 ($n = 6$) years and the adult group of 18–21 years ($n = 14$). The age at which the difference lost significance denoted the age when individuals approached adult performance.

For all statistical tests in the five games, the dependent measures were modeled as a linear combination of fixed and random effects and evaluated by a linear mixed model (LMM). For the choosing task, a binary response variable (right or wrong) was fitted with a binomial distribution with a *logit* link function and evaluated with a generalized linear mixed model (GLMM).

A first statistical analysis was conducted on all museum data (individual trials) for each task to evaluate whether performance changed as a function of age. This GLMM had age as a continuous fixed factor and the experimental condition (where present) as fixed factor. It also had subject number as random factor. The main experimental design compared different age groups and three experimental conditions in the catching and choosing tasks (number of trees and ball speeds, respectively). The GLMM compared the performance in each game with respect to fixed effects (age groups and tree/window), while accounting for the variability between participants (random effect: participants). Note this statistical analysis included all individual trials of all participants. To identify the model that best fit each variable, an iterative procedure was adopted to assess whether the inclusion of interaction and random effects was justified.^{84–87} The model that we considered to fit all variables, including all the possible factors, was the following:

$$Y_{ij} = (\beta_0 + S_{0i}) + (\beta_A + S_{Ai})A_j + (\beta_C + S_{Ci})C_j + \beta_{AC}A_jC_j + \varepsilon_{ij}$$

where Y was the dependent measure for each participant i and each trial j , A was the age factor, C was the experimental condition factor (from 1 to 3, only for catching and choosing tasks), β were the fixed-effects coefficients, S were the random-effects coefficients (intercept and slope) for participants, and ε_{ij} were the residuals. It is worth noting that the age factor was a categorical predictor and the levels depended on the age bins defined in each data set. The adult age bin was chosen as the reference level, and the other age groups were compared to it.

The GLMMs were run in R Studio (2022.02.3, with packages *nlme*, *lme4*, *lmerTest*). Other analyses were conducted using custom-made scripts and functions in MATLAB.

Lastly, we analyzed inter-task correlations between the five games in a pairwise fashion. Since each game exhibited a different performance distribution, the performance variables of all five games were rank-ordered (rank 1 being the best) before calculating Pearson correlation coefficients. This method was identical to how Spearman's correlation coefficients are calculated. We included participants who completed all five games at the museum ($n=219$). To eliminate the confound from age-related variability, we ran partial correlations with age as the controlling variable. The residuals were computed from the exponential fits as a function of age in each task. The inter-task correlations of the residuals were calculated in a pair-wise manner. The p values were Bonferroni corrected. The asterisks in Figure 7 indicate significant inter-task correlation pairs (* $p < 0.05$, ** $p < 0.005$, corrected).