Journal of Experimental Child Psychology 189 (2020) 104696

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Look before you fit: The real-time planning cascade in children and adults

Ori Ossmy, Danyang Han, Minxin Cheng, Brianna E. Kaplan, Karen E. Adolph *

Department of Psychology and Center for Neural Science, New York University, New York, NY 10003, USA

ARTICLE INFO

Article history:
Received 27 March 2019
Revised 31 July 2019
Available online 28 October 2019

Keywords:
Problem solving
Object fitting
Eye tracking
Motor development
Planning
Manual skills

ABSTRACT

Goal-directed actions involve problem solving—how to coordinate perception and action to get the job done. Whereas previous work focused on the ages at which children succeed in problem solving, we focused on how children solve motor problems in real time. We used object fitting as a model system to understand how perception and action unfold from moment to moment. Preschoolers ($N = 25$) and adults ($N = 24$) inserted three-dimensional objects into their corresponding openings in a "shape-sorting" box. We applied a new combination of real-time methods to the problem of object fitting—head-mounted eye tracking to record looking behaviors, video microcoding to record adjustments in object orientation between reach and insertion, and real-time analysis techniques (recurrent quantification analysis and Granger causality) to test the timing relations between visual and manual actions. Children, like adults, solved the problem successfully. However, adults outperformed children in terms of their speed of fitting, and speed depended on when adjustments of object orientation occurred. Adults adjusted object orientation during transport, whereas children adjusted object orientation after arriving at the box. Children's delays in adjustment resulted from delays in looking at the target shape and its corresponding aperture. Findings show that planning is a real-time cascade of perception and action, and looking provides the basis for planning actions prospectively. We suggest that developmental improvements in problem solving are driven by real-time changes in the instigation of the planning cascade and the timing of its components.

* Corresponding author.
E-mail address: karen.adolph@nyu.edu (K.E. Adolph).

https://doi.org/10.1016/j.jecp.2019.104696

Introduction

Everyday behavior is rife with motor problems—how to grasp the handle of a tool, open a latch, or navigate an obstacle (Adolph & Robinson, 2015; Bernstein, 1996; Keen, 2011). To solve such problems, adults typically use visual information to plan what actions to do and when and how to do them (Land & Hayhoe, 2001). The requisite visual information, however, does not normally arrive happenstance on passive receptors. Instead, gathering visual information is a deliberate motor act. We do not merely see; we *look*. As in the adage, ''Look before you leap," adults guide upcoming actions by intentionally moving their head and eyes to direct their gaze at relevant parts of the environment (Gibson, 1979). Children learn to do likewise (Franchak, Kretch, Soska, & Adolph, 2011; Kretch & Adolph, 2017; Rosander & von Hofsten, 2011; von Hofsten, 2003).

From visual information gathering to subsequent actions, solving everyday motor problems reflects a real-time "planning cascade"—a sequence of behaviors in which earlier events lay the foundation for later events (see Oakes & Rakison, 2019). Optimally, the cascade begins by looking at the task-relevant features and differentiating the appropriate visual information; the information is then used to formulate an action plan; and finally, the plan is translated via muscle activations into efficient movement. If any component in the cascade is delayed, is missing, or goes awry, the consequent action will be inefficient or will require online modifications—such as readjusting a grasp to wield a tool. In the worst case, the action fails and the problem remains unsolved—as when a latch remains locked or walkers fall while trying to navigate an obstacle.

In adults, the real-time planning cascade is widely evident. Head-mounted eye tracking shows that when adults walk over uneven terrain, they point their gaze a few steps ahead. When they spot an upcoming obstacle, adults adjust their step length and speed as they approach so that they can navigate the obstacle smoothly and efficiently (Matthis & Fajen, 2014; Patla, 1991). Similarly, adults plan multistep manual actions beginning with their first look. While preparing a cup of tea, for example, adults look at the appropriate object before their hand begins moving to the target (Land & Hayhoe, 2001). As their hand grasps the tea kettle, they are already looking ahead to the teacup.

Like adults, infants and young children also display evidence of planning. Head-mounted eye tracking shows that experienced walking infants direct their gaze to obstacles in their path. They modify their walking patterns as they approach a narrow bridge and while crossing the obstacle (Kretch & Adolph, 2017). Even at 10 months of age, infants obtain visual information to guide their manual actions. While placing a ball into a tube, they look ahead to the tube before their hand arrives at the target (Rosander & von Hofsten, 2011). In such locomotor and manual tasks, the demand for gathering visual information is paramount because bridge width and target distance change from trial to trial.

However, infants and young children frequently show deficits in tasks that appear effortless for adults. Children, for example, grasp spoons and hammers with a habitual overhand grip—regardless of the direction of the handle—resulting in impossibly awkward grips that require subsequent adjustments (Comalli et al., 2016; Keen, Lee, & Adolph, 2014; McCarty, Clifton, & Collard, 1999). Infants mistakenly grasp the bowl end of the spoon, resulting in a handful of applesauce (McCarty et al., 1999). In these cases, children's behavior resembles trial-and-error exploration rather than the outcome of a plan formulated before implementing the manual action (Adolph & Robinson, 2015).

One possible explanation is that children's deficits reflect a breakdown in the real-time planning cascade. Children may fail to gather the relevant visual information before acting (Lockman, Fears, & Jung, 2018); put another way, children leap before they look. A second possibility is that children do gather the relevant visual information sufficiently early to guide their actions, but they fail to process the information adequately. In other words, the problem is not a lack of information but rather what they do with the information they obtain (Comalli et al., 2016; Shutts, Örnkloo, von Hofsten, Keen, & Spelke, 2009). A third related possibility is that deficits result from children's lack of dexterity. Even when they obtain the relevant information before implementing the action and process it appropriately, children cannot rapidly and smoothly execute the required movements (Rosenbaum & Feghhi, 2019). To add to the complexity, these possibilities are not mutually exclusive. In principle, children could suffer from all three.

Object fitting and the planning cascade

Object fitting offers researchers an ideal way to study the real-time planning cascade because it involves an observable sequence of visual and manual actions (Lockman et al., 2018). For example, to fit an object into a "shape-sorter" toy (Fig. 1A), children must recognize the spatial relations between the appropriate object and aperture (i.e., outer contour of object must match inner contour of aperture), reach for and grasp the object, adjust the object's orientation in three dimensions as they transport it to the aperture, and finally insert it into the aperture (Fig. 1B).

Traditionally, researchers take a normative, outcome-oriented approach to object fitting by identifying the age at which children succeed in various fitting tasks (for a review, see Lockman et al., 2018). The most dramatic age-related changes occur between 14 and 60 months. Before 12 months of age, infants cannot fit any object into a corresponding aperture, although they appear to be interested in the task (Örnkloo & von Hofsten, 2007). Between 20 and 24 months, toddlers fit symmetrically shaped objects into their corresponding apertures; by 30 months, they do so for irregular and asymmetrically

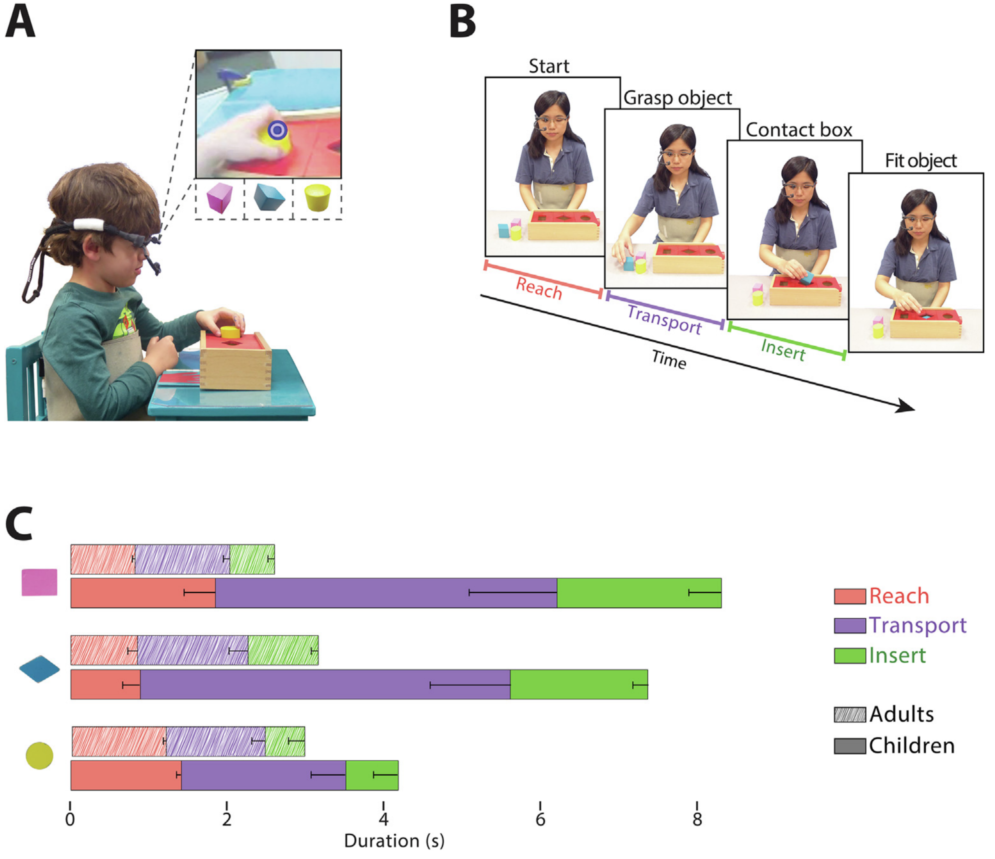

Fig. 1. Task description and speed of performance. (A) Object fitting task. Participants inserted three-dimensional shapes (rectangle, diamond, and circle) into corresponding apertures in a "shape-sorting" box while wearing a head-mounted eye tracker. The inset frame shows the view from the eye tracker (blue dot indicates gaze location) and an illustration of the three shapes. (B) Three-phase time course of fitting. Reach phase: Start of trial to grasp (participants grasped the object). Transport phase: Grasp to contacting the box with the object (participants transported the object from the table to the box). Insert phase: Box contact to fitting the object, which marked the end of the trial (participants fit the object into the aperture until it was fully inserted). (C) Average duration of each phase for each object. Adults performed faster than children, with significant differences during the transport and insert phases. Error bars denote standard errors.

shaped objects (Fragaszy, Kuroshima, & Stone, 2015; Örnkloo & von Hofsten, 2007; Street, James, Jones, & Smith, 2011). By 40 months, children solve the most complex forms of the problem, selecting among multiple irregular asymmetrical objects and/or multiple apertures for fitting (Örnkloo & von Hofsten, 2009; Shutts et al., 2009).

However, as Lockman et al. (2018) pointed out, the outcome-oriented approach addresses only the question of *when* children can solve a particular fitting problem. It does not provide insights into the process of *how* children solve the problem in real time—that is, how visual and manual actions in the planning cascade unfold from moment to moment to enable efficient fitting. One clue is that 14- and 18-month-olds succeed via clumsy trial-and-error matching of object shape to aperture shape (Örnkloo & von Hofsten, 2007; Shutts et al., 2009). They fumble around making adjustments to the object's orientation and position after already arriving at the appropriate aperture (Örnkloo & von Hofsten, 2007). In contrast, 30-month-olds adjust the orientation of the object as they transport it to the aperture. Thus, with age, children's adjustments occur at increasingly earlier points in the transport process, their adjustments become smoother, and they translate and rotate the object simultaneously rather than sequentially (Jung, Kahrs, & Lockman, 2015, 2018).

However, previous work on object fitting did not record children's direction of gaze. Thus, data on the real-time planning cascade from visual to manual actions is severely limited. Researchers could only infer whether and when children gathered visual information based on their manual actions. The extant data cannot distinguish whether the deficits observed in young children are due to lack of perceptual information, processing errors, poor manual dexterity, or a combination of factors. The current study addressed this important gap in the literature by using head-mounted eye tracking to record gaze locations.

The current study

Our overall goal was to use object fitting as a model system to characterize developmental changes in the real-time planning cascade of visual and manual actions. We bypassed the normative issue of whether younger and older children could fit objects into apertures by testing children at ages when they were sure to succeed, but potentially lacked adult-like coordination between visual and manual actions. Toward that end, we asked 3- to 5-year-old children and adults to fit circular, rectangular, and diamond-shaped blocks into corresponding apertures in a shape-sorter box. Participants wore a head-mounted eye tracker so that we could record their visual actions—whether and when they gathered visual information.

Based on previous work (Lockman et al., 2018), we expected adults to be more efficient than children in terms of the number of attempts before successful insertion of objects into their corresponding apertures and the time they needed to do it. We were particularly interested in where differences in visual and manual actions occurred in the time sequence—while reaching for the objects, transporting the objects to the box, and/or inserting the objects into the apertures. In principle, children's relative lack of dexterity could cause them to take longer than adults at each phase of object fitting. However, we hypothesized that manual dexterity is not the bottleneck, nor are deficits due to errors in processing the visual information. Rather, deficits in children's temporal efficiency result from failure or delay in obtaining the requisite visual information about objects and apertures during the reach and transport phases (during the insertion phase, visual information is relevant only for fine-tuning the object's orientation and position.) So, we coded the timing and duration of participants' looking fixations to the target objects and apertures during reach, transport, and insertion.

We also expected that adults would be more proactive and accurate in adjusting the object's orientation and position before it arrived at the aperture. That is, we expected adults to make adjustments earlier in the transport phase compared with children and to orient and align the objects more precisely prior to insertion. The rectangular and diamond-shaped blocks required alignment along both the vertical and horizontal axes, whereas the circle required alignment only along the vertical axis. Moreover, the dimensions of the rectangle created ambiguity about which way was up. So, we expected children to be most accurate in fitting the circle into its aperture and least accurate in fitting the rectangle.

In particular, we hypothesized that deficits in the timing and accuracy of adjustments are due to time lags in gathering visual information. We used recurrent quantification analysis (RQA) and Granger causality analysis to test this hypothesis. RQA allows researchers to determine which of two events occurs first when the events are reciprocal and repeated such as a heated conversation or, in this case, looking and adjusting. Granger causality analysis tests whether the occurrence of one event (looking) predicts the occurrence of a subsequent event (adjusting) when both events co-occur and are extended in time.

Method

Participants

We tested 25 children from 3.1 to 4.9 years of age (M = 4.0 years; 14 girls) and 24 adults from 18.8 to 26.5 years of age (M = 21.8 years; 15 women). All participants were healthy with normal vision. Children were recruited from advertisements, referrals, and a pool of families who expressed interest in participating in research when their children were born. Adults were recruited through word of mouth. Children received a robot toy, photograph magnet, and tote bag as souvenirs of participation, and adults received a photo magnet. With participants' permission, their videos and demographic data are shared in the Databrary web-based library (https://nyu.databrary.org/volume/434).

Head-mounted eye tracker

As in previous work (Franchak & Adolph, 2010; Franchak et al., 2011), participants wore a Positive Science head-mounted eye tracker (www.positivescience.com) so that we could record where they looked during the shape-fitting task (Fig. 1A). The hardware includes an ultra-light headgear, a transmitter, and a battery pack (total weight 375 g). The headgear was attached to an adjustable cap for stability and connected via wire leads to a computer on the other side of the room. A wide rubber belt (pictured in Fig. 1A) held the leads behind participants' back so that the leads did not interfere with participants' manual actions, and the transmitter and battery pack were attached to the belt.

The headgear consists of two small cameras mounted on child- or adult-sized eyeglass frames. The scene camera points out (54.4° × 42.2° field of view), and the eye camera points in toward the right eye (illuminated with an infrared emitting diode). Yarbus software synced videos from the two cameras, and PSEye Camera software allowed online monitoring of the eye and scene videos. To calibrate the tracker, seated participants fixated colored dots on a black display board on the table. We calibrated the data offline with Yarbus software by manually indicating the target locations on the appropriate video frames. The software calculates frame-by-frame point of gaze within the scene camera video (spatial accuracy ~ 2°) and produces a composite video with the point of gaze indicated on each video frame by a 4° radius circular cursor (pictured in Figs. 1A and 2A).

Procedure and shape sorter

To encourage children to wear the eye tracker, we told them that they were "science robots"—an appellation they received with delight. Participants sat in front of a child-sized table facing the experimenter. As shown in Fig. 1A, we asked participants to fit three differently shaped and colored wooden blocks into a wooden shape-sorter box: a yellow circle (4 cm high × 4.3 cm in diameter), a pink rectangle (4 cm high × 4 cm long × 3 cm wide), and a blue diamond (4 cm high × 4 cm long × 6 cm wide). The box (11 cm high × 27 cm long × 5.5 cm wide) had three apertures matched to the shapes of the three objects. The locations of the apertures were adjustable (right, middle, or left side of box) and were randomized across participants.

Participants received two trials, each with all three objects. The starting locations of the objects relative to the box (right, middle, or left) were randomized across participants and trials. To start each trial, the experimenter placed the shape-sorter box and three objects on the table and asked partici-

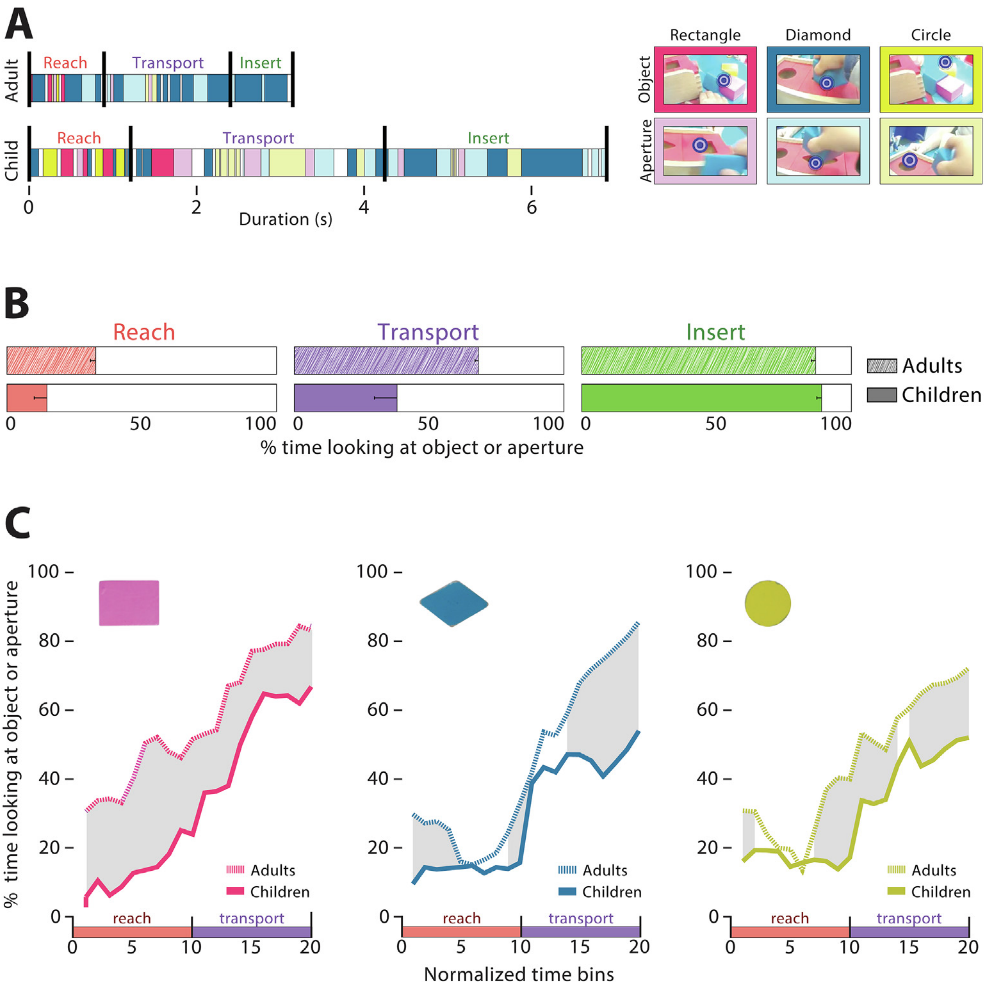

Fig. 2. Real-time looking patterns. (A) Real-time progression of looking during diamond fitting in one exemplar adult (top row) and one exemplar child (bottom row) across the three phases of the fitting task. Each color represents one of the six regions of interest (ROIs); see legend to the right of the timelines for an example. The adult consistently looked at the diamond-shaped object and its matched aperture from the midst of reaching, whereas the child looked primarily at the end of the transport phase. (B) Percentages of looking time at the object and its matched aperture while reaching, transporting, and inserting. Error bars denote standard errors. During reach and transport, adults fixated the object and its matched aperture more than children. (C) Average percentage times looking at the object and its matched aperture as a function of time bins during reach and transport. Gray areas denote significant differences between adults and children.

pants to fit the objects into the box in any order. Trials ended when all three objects were inserted into their matched apertures with no time restriction.

Sessions were videotaped at 30 frames per second from overhead and side views so that all manual actions were visible. The two views were mixed online onto a single video, which we then synced off-line with the eye-tracker videos for ease of later video coding and data processing. One trial from an adult and one trial from a child were excluded from analyses because of technical issues. The second trial from two children was excluded because they refused to complete the task. Thus, the total dataset comprised 94 trials and 282 object–aperture fits, 47 trials and 141 object-aperture fits per age group.

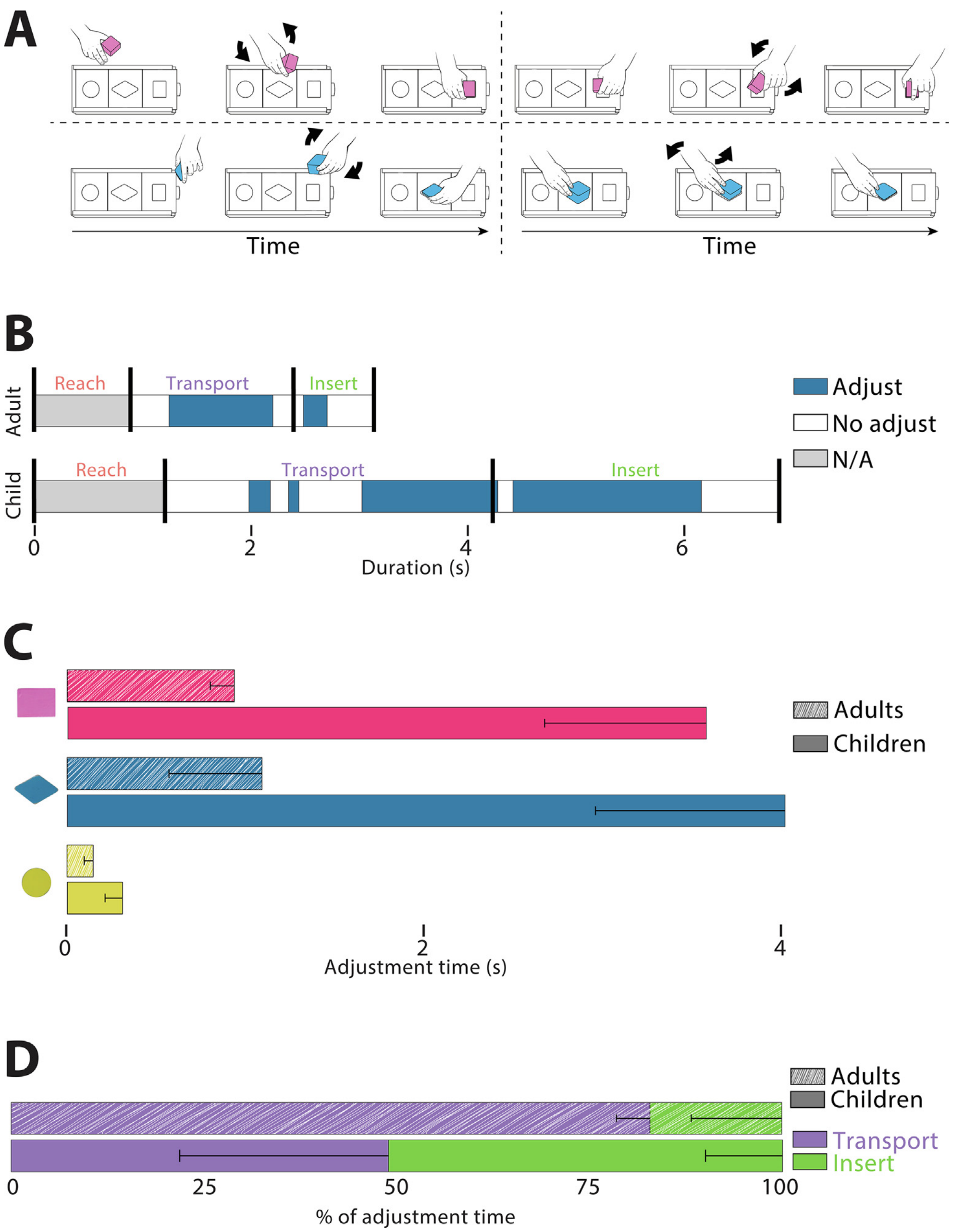

Fig. 3. Real-time adjustment patterns of object orientation and position. (A) Examples of adjustments during transport phase (left panel) and insert phase (right panel). (B) Real-time progression of adjustments during the same diamond fitting as in Fig. 2A. The adult initiated adjustment before the child, and most of the adjustment happened during the transport phase. Coders scored adjustments only after the reach phase when the object was grasped. (C) Average durations of adjustments accumulated during transport and insert per object. Overall, children were slower to adjust the shapes, especially the rectangle (pink) and the diamond (blue). (D) Ratios between phases in which participants changed the orientation of the object to match the aperture. Children changed orientation during the insert phase more than adults, whereas adults changed orientation during the transport phase before the object contacted the box.

Data coding

A primary coder scored videos frame by frame using Datavyu software (www.datavyu.org) to time lock user-defined events to their location in the video. The coder identified each attempt to fit an object into the box, divided attempts into sequential phases, and identified the timing and duration of specific visual and manual actions. To ensure interobserver reliability, a second coder independently scored 100% of attempts, fitting phases, and manual actions and 25% of visual actions (a smaller percentage because coding fixations is so laborious). Disagreements between coders were resolved by discussion.

Attempts to fit

In principle, participants could attempt to fit an object multiple times. So, coders considered each time the object arrived at an aperture to be an attempt. Attempts ended successfully when the object was fully inserted into its same-shaped aperture. On unsuccessful attempts, participants tried to insert the object into the wrong aperture, temporarily gave up and switched to a different object, or lifted the object from the box for at least 1 s. The first attempt began when the objects were available on the table. Subsequent attempts began when the preceding attempt ended. In other words, the last frame of the prior attempt was the first frame of the next attempt. Coders agreed on 96.02% of attempts and outcomes, *kappa* = .84, $p < .001$.

Fitting phases

Fitting consisted of a reach, transport, and insertion phase (Fig. 1B). The reach phase began when participants' hands were empty and the object(s) was available on the table and ended when participants grasped a block. The transport phase began with the grasp and ended when the block touched the box. The insertion phase began when the block touched the box and ended when it was fully inserted into its corresponding aperture or when children returned the object to the table without fitting it successfully. Note that participants could make multiple attempts to fit during a single insertion phase if they tried multiple apertures. If they did not replace the shape on the table, then the attempt had only transport and insert phases. Manual actions that did not lead to an attempt to fit (e.g., touching the objects or box) were excluded. Coders showed high agreement for the durations of fitting phases, $rs(47) > .95$, $ps < .001$.

Visual actions

Coders scored every fixation to each object and aperture (i.e., six regions of interest [ROIs]) based on the location of the gaze cursor in the eye-tracking video. The photographs in Figs. 1A and 2A show examples of fixations to objects and apertures. Participants could fixate the object in hand and its corresponding aperture and/or other objects and apertures, or they could look at irrelevant things (e.g., the table, the experimenter). The timelines in Fig. 2A show the moment-to-moment changes in fixations to objects and apertures for a typical adult (top) and a typical child (bottom) while they fit the diamond into the diamond-shaped aperture. Interobserver agreement for accumulated durations of fixations to each region was high, $rs(47) > .91$, $ps < .001$, and the two coders showed exact agreement for at least 88.72% of video frames for six ROIs, *kappas* > .81, $ps < .001$.

Adjustments of object orientation and position

The most interesting manual component of planning was the timing of participants' adjustments to object orientation (rotating the asymmetrical rectangle and diamond to fit the aperture along the horizontal axes and turning all three objects along the vertical axis) and horizontal position (aligning it with the corresponding aperture). So, coders scored the onset and offset of each adjustment during the transport and insertion phases (Fig. 3A). (Of course, participants could not adjust the orientation or position of the object until they grasped it, so adjustments were not possible during the reach phase.) For the transport phase, coders focused on adjustments in orientation because translations in position were indistinguishable from transporting the object to the box. For the insertion phase, coders focused on both adjustments in orientation and in horizontal position. Interobserver agreement for accumulated durations of adjustments was high, $rs(47) > .79$, $ps < .001$.

Accuracy of adjustments during transport

As in previous work, coders scored the accuracy of adjustments during transport—the object's rotational alignment, vertical alignment, and horizontal position in the first video frame after it arrived at the shape-sorter box (last frame of transport phase and first frame of insertion phase). The images at the top of Fig. 5 (see Results) depict correct and incorrect adjustments during transport. Rotational adjustments were considered correct if the object was within 15° of the appropriate orientation (based on the overhead camera view) and were considered incorrect if it was more than 15°. Vertical adjustments were correct if the object was within 15° of the appropriate orientation (based on the side camera view) and were incorrect if it was more than 15°. Horizontal position was correct if the object covered at least 70% of the same-shaped aperture and was incorrect if it was less than 70%. Coders compared each target video frame with a photograph of the object at the 15° or 70% criterion. The pre-

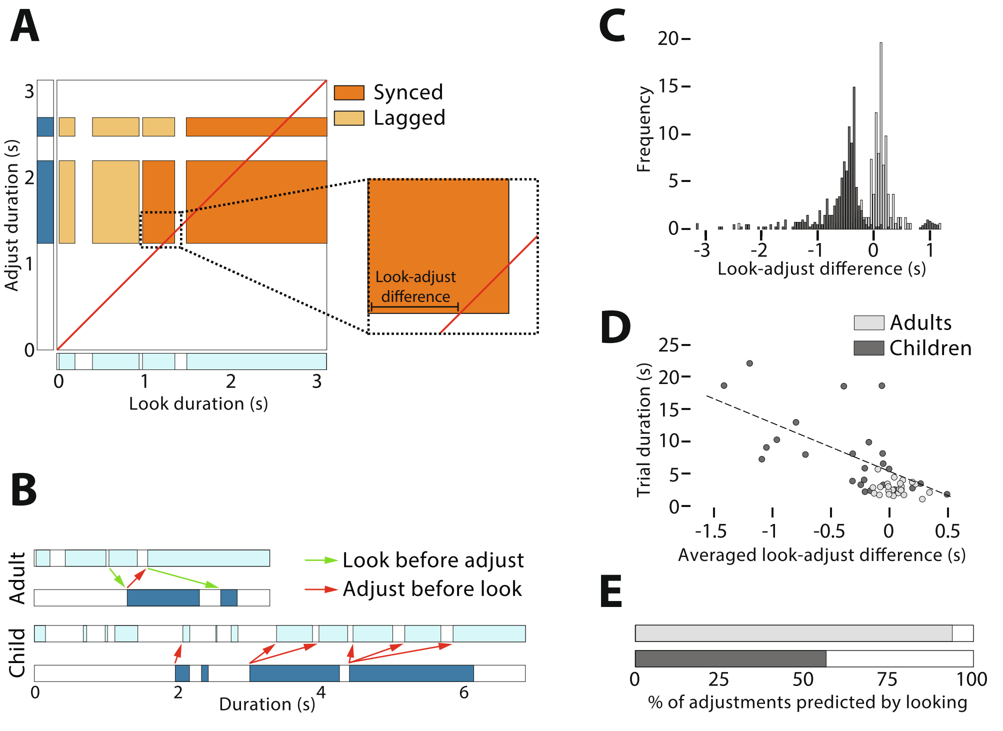

Fig. 4. Cascade from looking to adjusting. (A) An example recurrence plot of look–adjust coupling from one adult (same adult shown in Figs. 2A and 3A). The light and dark blue timelines along the *x* and *y* axes show a sequence of adjusting and looking during one trial. The red "real-time" diagonal line denotes the same moments for looking and adjusting. The recurrence of looking and adjusting is plotted as dark orange boxes along the red diagonal line (imagine "smearing" each box denoted in the timelines to the edges of the plot). The look–adjust difference is the time between the bottom-left corners of each dark orange box and the real-time diagonal (see Method). Light orange boxes (i.e., boxes not touching the red diagonal) show adjustment events that lagged so far behind looking events that the two events did not overlap. (B) Examples of a look–adjust coupling from one adult and one child (same participants pictured in Figs. 2A and 3A). The value of the look–adjust difference that was calculated with recurrent quantification analysis represents the temporal shifts that are marked with red and green arrows. For example, the adult had two recurrences in which looking started before adjusting and had one recurrence in which adjusting started before looking. According to the child's timeline, adjusting started prior to looking in all look–adjust recurrences. (C) Distribution of look–adjust differences. Adults (light gray) exhibited a small difference between the two events, and the distribution hovers around zero. Children (dark gray) showed mostly negative differences or positive differences that were higher than adults. (D) Relation between participants' average look–adjust difference and their full trial duration (i.e., fitting all three objects). Participants with larger negative values of the look–adjust difference score were slower to complete the task. (E) Using Granger causality, we tested whether looking could predict adjusting in fitting attempts when participants had positive look–adjust difference scores. Prediction was higher in adults (89.9%) compared with children (59.4%).

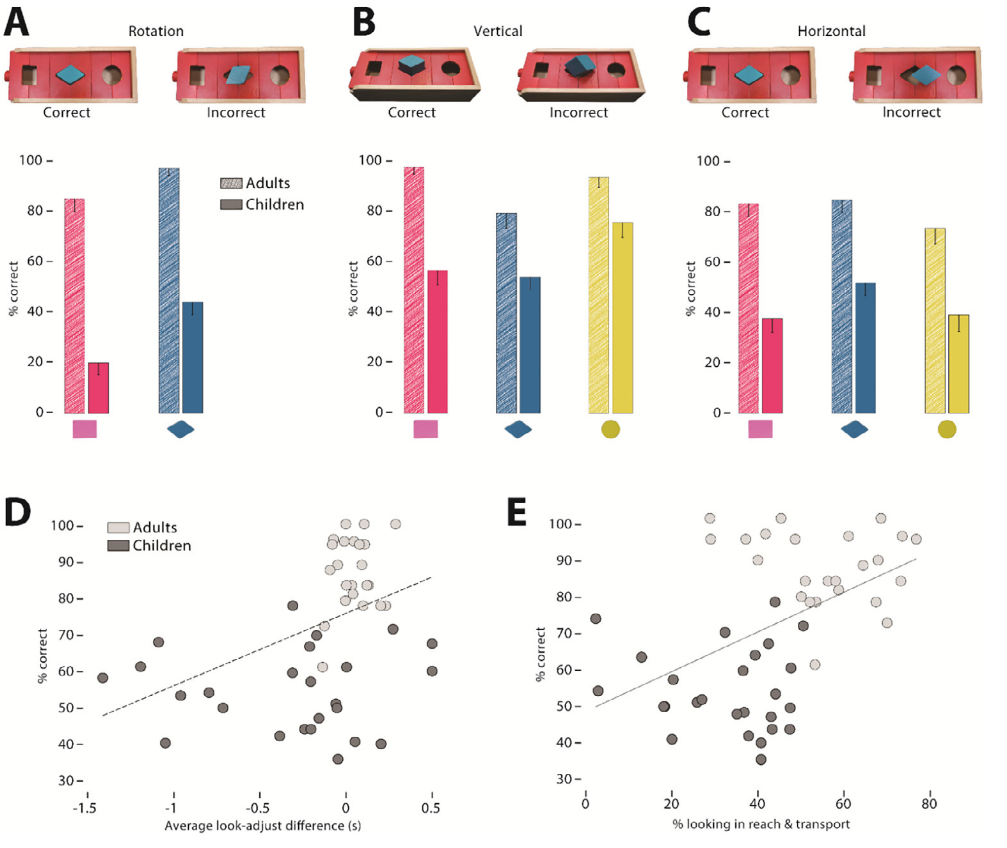

Fig. 5. Adjustment accuracies. (A–C) Accuracy in adjustments of object orientation. Top panel: Examples of correct and incorrect rotational, vertical, and horizontal adjustment. Bottom panel: Adults were more accurate in adjusting orientation regardless of shape and type of alignment. (D) Relation between participants' adjustment accuracy (averaged across shapes and type of alignment) and their look–adjust difference. Participants with positive differences were more accurate in adjusting object orientation. (E) Relation between participants' adjustment accuracy and their percentage of looking time at the object and its matched aperture prior to contact with the box. Participants who looked more were better in adjusting orientation.

cision criteria were based on Örnkloo and von Hofsten (2007). Coders agreed on the accuracy of rotational adjustment, vertical adjustment, and horizontal position for 98.64%, 97.82%, and 97.00% of insertions, respectively, *kappas* > .94, *ps* < .001.

Results

Preliminary analyses showed no effects of children's age or gender on their visual or manual actions, all *ps* > .22, so these factors were collapsed in subsequent analyses. Likewise, preliminary analyses showed no effects of fitting order on performance, *ps* > .80. Moreover, participants did not consistently fit the shapes in the same order across the two trials (only 7 participants used the same order in both trials). Thus, we averaged the data for each shape and participant across the two trials.

Attempts and duration of fitting phases

As we expected, every child and adult successfully fit every object into its corresponding aperture. However, as we also anticipated, children were less efficient than adults. Although unsuccessful attempts were rare (25/141 of object–aperture fits for children and 3/141 for adults), children averaged more attempts ($M = 1.64$, $SD = 0.65$) than adults ($M = 1.06$, $SD = 0.11$), $t(47) = 4.30$, $p < .001$. Only children tried to fit an object into the wrong-shaped aperture (10 unsuccessful attempts). For the other unsuccessful attempts, participants returned the block to the table or held it aloft. However, multiple attempts were relatively rare; a total of 15 children required multiple attempts, but 11 erred only once in one or two fits and only 4 children erred twice in a single fit.

Children's inefficiency was most evident in the time it took them to fit the objects. Children required a mean of 19.59 s ($SD = 11.96$) to complete each trial (fit all three objects), and adults required a mean of only 9.05 s ($SD = 2.21$), $t(47) = 4.24$, $p < .001$. As shown by the total length of the bars in Fig. 1C, children took longer for every shape compared with adults, and the slowdown was most evident for the rectangle and diamond compared with the circle. A 2 (age) $\times$ 3 (object) analysis of variance (ANOVA) on total time to fit the object confirmed a main effect for age, $F(1, 141) = 25.29$, $p < .01$, and an interaction between age and object, $F(2, 141) = 3.35$, $p < .04$. Sidak-corrected post hoc comparison tests for each age group showed that children were faster in fitting the circle compared with the other shapes, $ps < .05$.

Timing inefficiencies were not distributed equally across the three fitting phases. Instead, as shown in Fig. 1C, children were notably slower than adults in the transport and insertion phases for all three objects and during the reaching phase for the rectangular object. A 2 (age) $\times$ 3 (object) $\times$ 3 (phase) ANOVA on time confirmed main effects for age, $F(1, 423) = 34.76$, $p < .01$, and phase, $F(2, 423) = 17.11$, $p < .01$, as well as two-way interactions for age $\times$ object, $F(2, 423) = 4.04$, $p < .02$, and age $\times$ phase, $F(2, 423) = 7.23$, $p < .01$. The main effect of object and the object $\times$ phase interaction were nearly significant, $F(2, 423) = 2.98$, $p = .05$ and $F(4, 423) = 2.50$, $p = .05$, respectively. Sidak-corrected post hoc tests for each age group showed longer times for the transport and insertion phases in children, all $ps < .05$.

Real-time visual actions

The eye-tracking data generated detailed streams of looking to the six ROIs for each of the six object–aperture fits. Fig. 2A shows timelines of gaze data for a typical adult and child as they fit the diamond. As shown by the different colors, the adult maintained gaze on the diamond block (dark blue bars) and diamond-shaped aperture (light blue bars) for most of the sequence. Moreover, as shown by the light blue bars, the adult fixated the appropriate aperture during the reach phase and early in the transport phase. In contrast, the child spent less time looking at the diamond block and aperture during the reach and transport phases (denoted by the wild assortment of colored bars). The child did not point his gaze to the aperture until late in the transport phase and then switched between object and aperture during the insertion phase. As shown by the thin white bars, both the adult and child spent little time looking away from the objects and apertures. Across all participants, adults looked at the objects and apertures a mean of 89.97% of the time, and likewise children looked at the objects and apertures a mean of 85.12% of the time.

Because each object–aperture fit had a different duration, we transformed fixation times into the percentage of each phase in which participants looked at the relevant object and aperture. Both adults and children looked at the relevant object and/or aperture during the reach and transport phases for at least one frame. However, adults spent a larger percentage of the reach and transport phases looking at the relevant object and/or aperture than children (Fig. 2B). During the insertion phase, looking was similar in adults and children. A 2 (Age) $\times$ 3 (Phase) ANOVA confirmed main effects for age, $F(1, 141) = 25.65$, $p < .01$, and phase, $F(2, 141) = 37.02$, $p < .01$, as well as an Age $\times$ Phase interaction, $F(2, 141) = 5.31$, $p < .01$. Sidak-corrected post hoc tests confirmed more looking by adults during the reach and transport phases, all $ps < .05$.

To examine moment-to-moment changes in visual actions across each phase, we split each phase into 10 time bins and calculated the percentage of each bin in which participants fixated the relevant object and/or aperture. We compared looking in children and adults at each time bin, using *t* tests with a Bonferroni correction for multiple comparisons (alpha set to .001).

Fig. 2C depicts looking across the reach and transport phases. For all three objects, compared with children, adults looked more at the relevant object and/or aperture earlier during the reach and transport phases. Adults started by fixating the object they intended to grasp, scanned the scene for the relevant aperture in the midst of reaching (drop in looking at task-relevant ROIs in the gray area), and then fixated the task-relevant ROIs again to change the orientation properly before making physical contact with the box. Earlier looking in adults during both phases is clearly evident for the rectangle, as shown by the gray area (denoting significant differences). When children fit the rectangle, they started looking at the ROIs while reaching, but to a lesser extent compared with adults.

Earlier looking in adults was also evident for the diamond and circle, although the overall pattern differed. As shown by the gray regions, adults looked at the object/aperture early in the reach phase, then looked away from the ROIs, and then began looking again prior to transport and continually throughout transport. In contrast, when children fit the circle or diamond, they barely looked at the object or aperture during reaching (<20% for both objects) and showed a moderate increase in looking during transport. Taken together, children's real-time looking patterns suggest that they do not generate the relevant visual information ahead of time and resort to tactile information to guide their successful fitting.

Real-time adjustments of object orientation

Adjustments of object orientation played out in real time. Fig. 3B shows the moment-to-moment adjustments performed by the same adult and child shown in Fig. 2A. As shown in Fig. 3C, overall, children spent more time adjusting object orientation ($M = 2.75 \pm 1.2$ s) compared with adults ($M = 0.75 \pm 0.3$ s), $t(47) = 5.10$, $p < .001$. Fig. 3C also shows that the differences in adjustment times between children and adults derived mainly from fitting the rectangle and diamond but not the circle.

In addition, children adjusted object orientation mostly toward the end of the trial; a mean of 51% of their adjustment time was during the insert phase compared with a mean of 17% in adults, $t(47) = 4.90$, $p < .001$ (Fig. 3D). Children's real-time action patterns point to a slowdown in adjusting object orientation. Compared with adults, children began adjusting too late in the game, made fewer adjustments prior to fitting, and thus required more time to insert the objects.

Looking and adjusting: Recurrent quantification analysis

We used RQA to quantify the dynamics between participants' looking patterns and their adjustments of object orientation (for another example of RQA applied to looking patterns, see Yu & Smith, 2013). Our analyses focused only on the transport and insert phases because object orientation cannot be adjusted during reaching.

Fig. 4A depicts the method for one adult fitting one object. We aligned the looking time series (x axis) and the adjustment time series (y axis). The x axis represents times the participant looked at the target object or its matched aperture. The y axis represents times the participant adjusted the object. Regions touching the red diagonal indicate simultaneous looking and adjusting. Any part of an orange block not touching the diagonal denotes asynchrony between looking and adjusting; orange regions above the diagonal represent looking preceding adjustments, and orange regions below the diagonal represent adjustments followed by looking. The blocks touching the red diagonal are colored dark orange to highlight that only these blocks were used in the analysis. White regions represent no looking or adjusting.

The structure of the joint recurrence plot allowed us to quantify the pattern of coordination between looking and adjusting by calculating time differences indicated by the sequence of dark orange blocks. Latencies between looking and adjusting are the distances between the left-bottom corner of the dark orange blocks and the diagonal (see inset in Fig. 4A). A positive time difference indicates that looking preceded adjustment. A negative time difference indicates that participants adjusted before looking. Finally, we calculated the distribution of latencies for children and adults.

As shown in Fig. 4C, time differences between looking and adjusting were distributed differently in children and adults (Kolmogorov–Smirnov test, $p < .001$). In adults, adjustments were more likely to follow looking ($M_{\text{look-adjust}} = +0.2$ s), whereas children initiated adjustments before looking at the object or its matched aperture ($M_{\text{look-adjust}} = -0.87$ s). Crucially, fitting durations were negatively correlated with the time difference between looking and adjusting, $r(47) = -.57$, $p < .001$ (Fig. 4D).

Moreover, the negative correlation held for children alone, $r(23) = -.42$, $p < .05$. That is, children for whom looking preceded adjustment were faster to fit the object. In contrast, children for whom looking followed adjustments were slower to fit. The look-adjust difference was not correlated with the number of fitting attempts because multiple fitting attempts resulted primarily from poor adjustments during the insert phase, $r(47) = -.16$, $p = .27$.

Looking predicts adjustments: Granger causality

To investigate the nature of coordination between participants' looking patterns and their adjustments of object orientation, we used Granger causality analysis to determine whether gaze locations predict subsequent adjustments in object orientation (Granger, 1969; Kamiński, Ding, Truccolo, & Bressler, 2001). Granger causality is a statistical concept of causality for testing whether one time series is useful in forecasting another. Causality was calculated between the looking time series (*x* axis in Fig. 4A) and the adjustment time series (*y* axis in Fig. 4A). If accuracy in predicting adjustment times from looking times is significant and exceeds the accuracy in predicting looking times from adjustment times, we can conclude that gaze locations "Granger cause" changes in orientation. When causality cannot be determined because looking times did not predict adjustment times, or both looking and adjustment times are equally predictive of the other, we cannot consider looking as a causal factor in changing orientation. Because the time series of looking and adjusting are categorical, we convolved these time series with a small Gaussian kernel and calculated Granger causality on the convolved time series (Sameshima & Baccalá, 1999). The lag length selection (i.e., the sliding window used to determine change in orientation based on looking) was chosen using the Bayesian information criterion—a criterion used to choose models for particular datasets (Bishop, 1995). We used 500 ms as a lag length for the algorithm.

Importantly, we tested Granger causality only when the time difference between looking and adjusting was positive (i.e., when participants looked at the object or its matched aperture before adjusting the object). Of course, if looking occurred only after adjusting, gaze locations could not predict adjustments in object orientation.

The analysis showed that looking was more likely to *predict* adjustments in adults compared with children (Fig. 4E). For a mean of 89.9% of their object fittings, adults' gaze location reliably predicted when they adjusted the object to the aperture (alpha < .05) compared with a mean of 59.4% of children's fittings.

Accuracy of adjustments and looking

Finally, adults and children differed in the accuracy of their adjustments (as measured when the block arrived at the box). Adjustments were either accurate or inaccurate (Fig. 5A–C, top panel). The bottom panel of Fig. 5A–C shows that adults had more accurate adjustments than children in rotational alignment, $t(47) = 9.10$, $p < .001$, vertical alignment, $t(47) = 4.20$, $p < .001$, and horizontal alignment, $t(47) = 5.90$, $p < .001$. (The circular object did not require rotational alignment.) Children were better at adjusting vertical orientation ($M = 61.5 \pm 5.3\%$ accurate adjustments) than horizontal ($M = 42.7 \pm 4.6\%$) or rotational ($M = 31.25 \pm 4.0\%$) orientation, $F(2, 199) = 8.39$, $p < .001$.

With both children and adults included, the average accuracy in adjustments was correlated with the look–adjust difference, $r(47) = .33$, $p < .03$ (Fig. 5E) and with the percentage of looking time during reach and transport, $r(47) = .38$, $p < .01$ (Fig. 5D). However, these correlations did not hold when we analyzed only children because their look–adjust difference scores were largely negative and they had low looking times during reach and transport.

Discussion

We used object fitting as a model system to understand how visual and manual actions unfold from moment to moment while children and adults solve motor problems. A remarkable aspect of proficient problem solving is the rapidity of achieving a goal. In just a few moments, perceptual, cognitive, and motor systems cooperate to produce an optimal solution (Prinz, 1997; Rosenbaum, Chapman, Weigelt, Weiss, & van der Wel, 2012; Wolpert, Diedrichsen, & Flanagan, 2011). For example, each adult in our study contributed six trials of object fitting in only 20 s or so of data. Each child did likewise in about 60 s of data. But real-time analyses of these brief snippets of behavior provided detailed insights into the contributions of perception, cognition, and action for planning. In less than 3 s, adults detected object size and shape, recognized the spatial relations between object and aperture, reached for and grasped the object based on its spatial structure, aligned the object to the aperture while transporting it to the shape-sorter box, and finally inserted the object. In such a rapid process, actions must be precisely timed and sequenced, beginning with visual actions to gather task-relevant information followed by executing the appropriate manual actions (von Hofsten, 2003).

As we expected, children, like adults, solved the problem of fitting objects into their same-shaped apertures (Lockman et al., 2018). However, adults outperformed children in terms of their speed in fitting, and speed depended on the timing of adjustments in object orientation. Adults adjusted object orientation while transporting objects to the shape-sorter box, whereas children adjusted object orientation after arriving at the box. It was as if children were "waiting" for tactile information to guide their solutions. Why?

We tested three hypotheses regarding the deficits in children's planning: that deficits result from a failure to obtain sufficient visual information early enough to form an efficient plan, that processing errors after obtaining the visual information interfere with planning, and that lack of manual dexterity hinders execution. To do so, we applied a new combination of real-time technologies to the problem of object fitting: head-mounted eye tracking to record children's visual actions from moment to moment, frame-by-frame video coding to determine adjustments in object orientation between reach and insertion, and real-time analysis techniques (RQA and Granger causality) to test the timing relations between visual and manual actions.

Our approach points to delays in visual actions—deliberate gathering of visual information—as a key cause of children's inefficient object adjustments. Real-time analyses showed that children's delay in adjustment was due largely to a delay in looking at the target shape and its corresponding aperture and an insufficient amount of looking when they did so. The findings support the hypothesis that deficits in children's planning result from timing delays in a real-time planning cascade in which visual actions (looking at the object and aperture) must precede manual actions (adjusting the object's orientation to the aperture). In short, poor planning results from a breakdown in the real-time cascade.

Although results indicate that children are poor problem solvers because they fail to gather crucial visual information before acting, the findings do not exclude the other hypotheses. In principle, children might also be inefficient in processing the visual information they obtain, and/or their lack of manual dexterity might interfere with execution (e.g., diamonds and rectangles are more difficult to pick up than circles). The current study paves the way for future research to consider and rule out the other possibilities, for example, by using eye tracking with electroencephalography (to examine information processing) and motion tracking (to examine dexterity) simultaneously.

Planning in motor problem solving

In principle, the entire sequence of visual and manual actions can be planned ahead of time before implementing a solution. Indeed, planning ahead leads to more efficient solutions. Failure to plan in an earlier step can necessitate modifications in a subsequent step and, thus, can incur cost in terms of effort and time (Keen, 2011). So, an important challenge for solving motor problems concerns when to begin planning the sequence of actions—in the case of object fitting, when to look relative to reaching and adjusting object orientation.

Our findings expand on previous work that pointed to poor planning in children's object fitting—even after children acquire the ability to eventually insert shapes into their corresponding apertures. As in the current study, previous work with toddlers and preschoolers showed that objects were misaligned at the moment the object arrived at the aperture. The requirement for subsequent adjustments provided evidence of poor planning during reach and transport (Meyer, 1940; Örnkloo & von Hofsten, 2007, 2009). That is, even after children can recognize which shape belongs in which hole, they fail to plan ahead of time for the object's orientation. Recent work used high-speed motion tracking to record the real-time movements from reach to insertion (Jung et al., 2015, 2018). Before about 18 months of age, children adjust object orientation in two discrete steps: first a translational movement to transport the object to the aperture, followed by a rotational movement to align the object's orientation appropriately. After about 30 months, children combine translational and rotational movements into a single step during transport.

Use of head-mounted eye tracking in the current study revealed where and when children direct their visual attention. Adjusting object orientation during the transport phase yields faster fitting than adjusting the orientation during the insert phase. But most of children's looking at the object and corresponding aperture occurred during the insertion phase—too late to adjust object orientation during transport. Nonetheless, children did obtain some visual information about the object and the aperture before insertion (see Fig. 2), yet their performance was still poor. Presumably, children's looking cannot guide adjustments of object orientation during transport because children looked too little and too late. Accordingly, several studies showed that more complex visual information (e.g., variations in size or geometric shape of the object) causes more failures in object fitting (Shutts et al., 2009) and more deficits in adjustments to object orientation during object transport (Fragaszy & Cummins-Sebree, 2005; Fragaszy et al., 2015; Örnkloo & von Hofsten, 2007).

Planning to plan: The role of visual actions

We offer a final suggestion. Children's failure to obtain sufficient visual information at the start of the trial may reflect a failure of "planning to plan." It is possible that children do not realize that they need to deliberately point their eyes at the object and aperture to form an efficient plan for the manual action. Looking is itself an action that requires planning. Thus, the developmental shift from children's trial-and-error strategies to adults' efficient planning-ahead strategies might result from improvements in planning to plan—planning to get the perceptual information that is necessary for subsequent planning.

Conclusions

Our findings show that planning is a cascade of necessary visual and manual actions, and in so doing we suggest that looking must be planned as the first stage of the planning cascade. If the motor system is not used effectively during planning to sample the perceptual world and extract task-relevant information, then subsequent stages of planning will go awry. Our novel combination of eye tracking, video microcoding, and real-time analyses suggests that developmental changes in problem solving are driven by changes in the onset of the planning cascade and the moment of coupling between its components. The bottom line: Look before you leap.

Acknowledgments

This research was supported by grants to Karen Adolph from the National Science Foundation (NSF)/Social, Behavioral, and Economic Sciences (SBE)—U.S.–Israel Binational Science Foundation (BSF) (1627993), the National Institute of Child Health and Human Development (NICHD) (R01HD086034), and the Defense Advanced Research Projects Agency (DARPA) (N66001-19-2-4035). We are grateful to Anna Hyeji Shin and Nyla Elise Boynton-Jarrett for their help with data coding. The authors have no competing interests to declare.

References

Adolph, K. E., & Robinson, S. R. (2015). Motor development. In L. Liben & U. Muller (Eds.). *Handbook of child psychology and developmental science* (Vol. 2, pp. 114–157). New York: John Wiley.

Bernstein, N. A. (1996). On dexterity and its development. In M. L. Latash & M. T. Turvey (Eds.), *Dexterity and its development* (pp. 3–244). Mahwah, NJ: Lawrence Erlbaum.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Clarendon.

Comalli, D. M., Keen, R., Abraham, E., Foo, V. J., Lee, M. H., & Adolph, K. E. (2016). The development of tool use: Planning for end-state comfort. *Developmental Psychology, 52*, 1878–1892.
Fragaszy, D. M., & Cummins-Sebree, S. E. (2005). Relational spatial reasoning by a nonhuman: The example of capuchin monkeys. *Behavioral and Cognitive Neuroscience Reviews, 4*, 282–306.
Fragaszy, D. M., Kuroshima, H., & Stone, B. W. (2015). "Vision for action" in young children aligning multi-featured objects: Development and comparison with nonhuman primates. *PLoS One, 10*(10), e140033.
Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research, 50*, 2766–2774.
Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development, 82*, 1738–1750.
Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*, 424–438.
Jung, W. P., Kahrs, B. A., & Lockman, J. J. (2015). Manual action, fitting, and spatial planning: Relating objects by young children. *Cognition, 134*, 128–139.
Jung, W. P., Kahrs, B. A., & Lockman, J. J. (2018). Fitting handled objects into apertures by 17- to 36-month-old children: The dynamics of spatial coordination. *Developmental Psychology, 54*, 228–239.
Kamiński, M., Ding, M., Truccolo, W. A., & Bressler, S. L. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics, 85*, 145–157.
Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual Review of Psychology, 62*, 1–21.
Keen, R., Lee, M. H., & Adolph, K. E. (2014). Planning an action: A developmental progression in tool use. *Ecological Psychology, 26*, 96–108.
Kretch, K. S., & Adolph, K. E. (2017). The organization of exploratory behaviors in infant locomotor planning. *Developmental Science, 20*, e12421.
Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559–3565.
Lockman, J. J., Fears, N. E., & Jung, W. P. (2018). The development of object fitting: The dynamics of spatial coordination. *Advances in Child Development and Behavior, 55*, 31–72.
Matthis, J. S., & Fajen, B. R. (2014). Visual control of foot placement when walking over complex terrain. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 106–115.
McCarty, M. E., Clifton, R. K., & Collard, R. R. (1999). Problem solving in infancy: The emergence of an action plan. *Developmental Psychology, 35*, 1091–1101.
Meyer, E. (1940). Comprehension of spatial relations in preschool children. *The Pedagogical Seminary and Journal of Genetic Psychology, 57*, 119–151.
Oakes, L. M., & Rakison, D. H. (2019). *Developmental cascades: Building the infant mind*. Oxford, UK: Oxford University Press.
Örnkloo, H., & von Hofsten, C. (2007). Fitting objects into holes: On the development of spatial cognition. *Developmental Psychology, 43*, 404–416.
Örnkloo, H., & von Hofsten, C. (2009). Young children's ability to solve spatial problems involving a choice. *European Journal of Developmental Psychology, 6*, 685–704.
Patla, A. E. (1991). Visual control of human locomotion. In A. E. Patla (Ed.), *Adaptability of human gait* (pp. 55–97). North-Holland, Netherlands: Elsevier Science.
Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9*, 129–154.
Rosander, K., & von Hofsten, C. (2011). Predictive gaze shifts elicited during observed and performed actions in 10-month-old infants and adults. *Neuropsychologia, 49*, 2911–2917.
Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., & van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological Bulletin, 138*, 924–946.
Rosenbaum, D. A., & Feghhi, I. (2019). The time for action is at hand. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-018-01647-7. Advance online publication.
Sameshima, K., & Baccalá, L. A. (1999). Using partial directed coherence to describe neuronal ensemble interactions. *Journal of Neuroscience Methods, 94*, 93–103.
Shutts, K., Örnkloo, H., von Hofsten, C., Keen, R., & Spelke, E. S. (2009). Young children's representations of spatial and functional relations between objects. *Child Development, 80*, 1612–1627.
Street, S. Y., James, K. H., Jones, S. S., & Smith, L. B. (2011). Vision for action in toddlers: The posting task. *Child Development, 82*, 2083–2094.
von Hofsten, C. (2003). On the development of perception and action. In K. J. Connolly & J. Valsiner (Eds.), *Handbook of developmental psychology* (pp. 114–140). London: Sage.
Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience, 12*, 739–751.
Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye–hand coordination. *PLoS One, 8*(11), e79659.