Euclid's Random Walk:
Developmental changes in the use of simulation for geometric reasoning

Yuval Hart, L. Mahadevan, & Moira R. Dillon

Supplemental Materials and Methods

Preregistered exclusion criteria for the localization task

In the localization task, participants were asked to position a dot on the location of a triangle's missing vertex by clicking a mouse cursor. Seven isosceles triangles were presented, which had seven different triangle side-length values combined with two angle sizes and four base lengths (see **Table 2**). This setup allowed us to calculate the standard deviation in the y direction of participants' responses for each triangle size and then calculate the power law by which the standard deviations scaled with triangle side-lengths, $\sigma \propto L^n$. This calculation yielded the scaling exponent.

An underlying assumption of the scaling exponent is that the standard deviation of the responses for each triangle for each participant is related to the length of the triangle's side (with greater deviation for longer lengths). A lack of relation might occur when a participant's localization estimates differ greatly for the same triangle, perhaps because of large deviations from the true vertex location. In attempt to ensure that a meaningful scaling exponent was calculated for each participant, we specified in advance that participants who did not show a certain value and significance level in this relation would be excluded from the analysis. We set correlation and significance values to Pearson $r \geq 0.75$ and $p < .005$ (see **Fig. S1**). As described in the main text, these values turned out to be too restrictive, and below we repeat our analyses for the children excluded because of this criterion.

Supplemental Results

Findings from children excluded for their responses on the localization task

As described in the main text, our preregistered criteria led to the exclusion of 52 children based on the relation between the standard deviation of their localizations for each triangle and the length of each triangle's side on the localization task. To test the generalizability of the findings we report in the main text, we repeat our analyses with these excluded children.

Reasoning Task

First, as in the main analysis, children responded more accurately in the reasoning task on questions about the position versus angle size of the fragmented triangle's missing corner (P = 0.795, 95% CI = [0.738, 0.842], $p < .001$) and when there was a transformation to the angle sizes versus the distance between the two given corners (P = 0.761, 95% CI = [0.698, 0.814], $p < .001$). Neither the size of the transformation nor the implied area of the fragmented triangle presented with each question affected children's accuracy (ps $> .158$). Older children, moreover, were more accurate on this task than younger children (P = 0.546, 95% CI = [0.516, 0.575], $p = 0.002$).

Localization Task

As in the main analysis, the error with which children localized a triangle's missing vertex in the localization task grew significantly as triangle side-length grew ($p < .001$), but, unlike the main analysis, error did not grow less in older versus younger children ($p = .529$). This difference is likely due to the excluded children's noisier responses on this task. Finally, consistent with the main analysis, the relation between scaling exponent and age was not significant ($p = .550$).

Relation between simulation and reasoning

Across the sample of excluded children split by age at 10 years (≥ 10 years, N = 16; <10 years, N = 36), a binomial mixed-model logistic regression predicting reasoning accuracy by scaling exponent, age, and their interaction found no significant effect of scaling exponent ($P = 0.493$, 95% CI = [0.232, 0.758], $p = .964$) or age ($P = 0.279$, 95% CI = [0.074, 0652], $p = .237$). These findings were further characterized by a scaling exponent by age interaction ($P = 0.911$, 95% CI = [0.546, 0.989], $p = .033$). Individual contrasts revealed no relation between scaling exponent and reasoning for younger children ($P = 0.494$, 95% CI = [0.271, 0.719], $p = .958$), but a significant relation for older children ($P = 0.930$, 95% CI = [0.507, 0.994], $p =.048$). All of these results are consistent with the main analysis.

Gender as an additional predictor variable

Because we observed a significant effect of gender on children's performance on the reasoning task in the main sample of children, here we repeat our main analyses (N = 125) with gender included as an additional predictor variable.

Reasoning Task

A binomial mixed-model logistic regression evaluated the role of question type, transformation, size of the transformation, the two-way interactions between these variables, the implied area of the fragmented triangle, age, and gender on children's accuracy. As in the main analysis, children were more accurate on questions about the location versus angle size of the missing corner (P =.747., 95% CI = [0.673, 0.809], $p < .001$) and when there was a transformation to the angle sizes versus the distance between the two given corners (P =.716, 95% CI = [0.637, 0.784], $p <.001$). Children were also more accurate when they were asked about the location versus angle size of the missing corner after a distance transformation to the two given corners (P =.691, 95% CI = [0.596, 0.772], $p < .001$). The size of the transformation as well as the implied area of the fragmented triangle presented with each question did not affect children's accuracy (ps $> .479$). Finally, older children were more accurate on this task than younger children (P =.538, 95% CI = [0.507, 0.568], $p =.015$).

Localization Task

As in the main analysis, error grew less in older versus younger children after correcting for gender ($p = .039$), and the relation between scaling exponent and age was not significant after correcting for gender ($p = .665$).

Relation between simulation and reasoning

A binomial mixed-model logistic regression across the sample of children and correcting for gender showed that the relation between reasoning accuracy and scaling exponent was not significant ($P = 0.289$, 95% CI = [0.107, 0.581], $p = .151$).

The analysis with children below and above 10 years of age reveled results consistent with the main analysis. There was no significant effect of scaling exponent on reasoning accuracy (*P* = 0.497, 95% CI = [0.181, 0.816], *p* = *.987*), but a significant effect of age (*P* = 0.911, 95% CI = [0.586, 0.987], *p* = *.021*). These results were further characterized by the scaling exponent by age interaction (*P* = 0.104, 95% CI = [0.011, 0.548], *p* = *.072*). Consistent with the main analysis, individual contrasts revealed no relation between scaling exponent and reasoning for younger children (*P* = 0.497, 95% CI = [0.180, 0.816], *p* = *.987*), but a significant relation for older children (*P* = 0.103, 95% CI = [0.018, 0.412], *p* = *.019*).

Age-dependent differences in the relation between scaling exponent and reasoning based on a change-point analysis of reasoning-task accuracy

As outlined in the main text, a change-point analysis on children's accuracy on the reasoning task, with age binned by month and using a binary segmentation method (Scott & Knott, 1974) with a Bayesian information criterion (BIC) penalty type, revealed one change point at 10 years 3 months (**Fig. S3**). In a mixed-model binomial logistic regression using this age split (≥ 10 years 3 months, N = 59; < 10 years 3 months, N = 66) and predicting reasoning accuracy by scaling exponent, age, and their interaction, we find no overall effect of scaling exponent (*P* = 0.512, 95% CI = [0.193, 0.822], *p* = .949), but a significant effect of age (*P* = 0.937, 95% CI = [0.668, 0.991], *p* = .008). These results were further characterized by a scaling exponent by age interaction (*P* = 0.069, 95% CI = [0.007, 0.446], *p* = .033). Individual contrasts revealed no relation between scaling exponent and reasoning for younger children (*P* = 0.492, 95% CI = [0.176, 0.815], *p* = .966), but a significant relation for older children (*P* = 0.091, 95% CI = [0.016, 0.381], *p* = .013). All of these results are consistent with the main analysis.

Age-dependent differences in the accuracy of the angle questions alone

We planned an additional analysis of children's reasoning responses on the angle questions alone. We found that both the type of transformation and age significantly affected reasoning accuracy on these questions (Transformation: *P* = 0.742, 95% CI= [0.651, 0.816], *p* < *.001*; Age: *P* = 0.550, 95% CI= [0.510, 0.589], *p* = *.015*): Children were more accurate when the angles sizes versus the distance between the corners were changed; and older children were more accurate than younger children. The effects of the size of the transformation, the implied area, and the interaction between the type and size of the transformation were not significant (all *ps* > *.220*).

Finally, a binomial mixed-model logistic regression examining the relation between children's angle question accuracy and scaling exponent, age, and their interaction revealed no effect for age, scaling exponent, or their interaction (Scaling Exponent: *P* = 0.997, 95% CI = [0, 1], *p* = *.636*; Age: *P* = 0.644, 95% CI = [0.379, 0.843], *p* = *.285*; Scaling Exponent *Age: *P* = 0.382, 95% CI = [0.145, 0.694], *p* = *.469*).

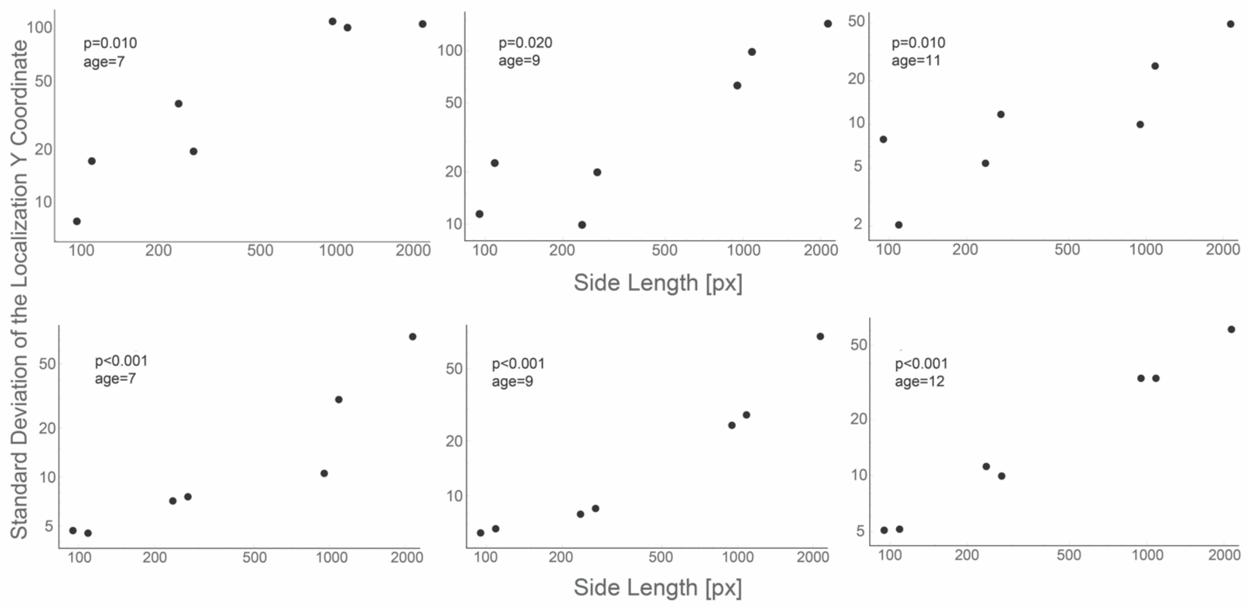
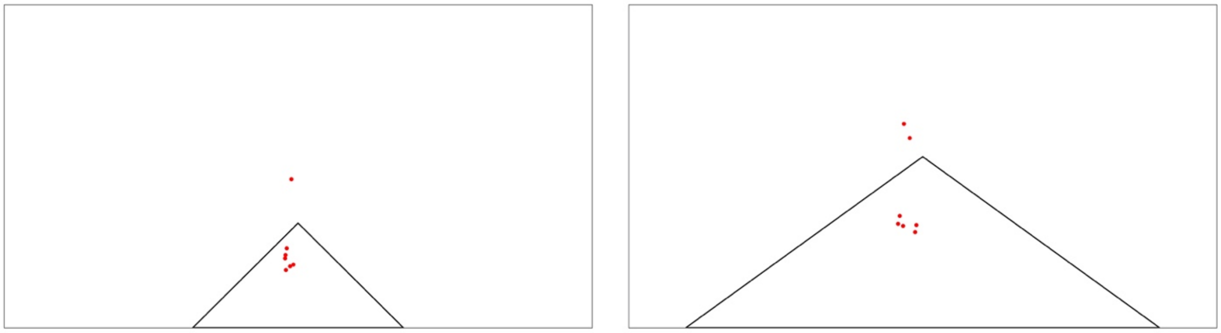


Fig. S1. Example plots of the relation between the standard deviations of participants' localization estimates in the y direction and the triangles' side lengths. **Top**. Examples of individual participants where the Pearson *r* correlation and significance value fell below the preregistered criteria. **Bottom**. Examples of individual participants where correlation and significance values fell at or above the preregistered criteria.

age=9, excluded child



age=9, included child

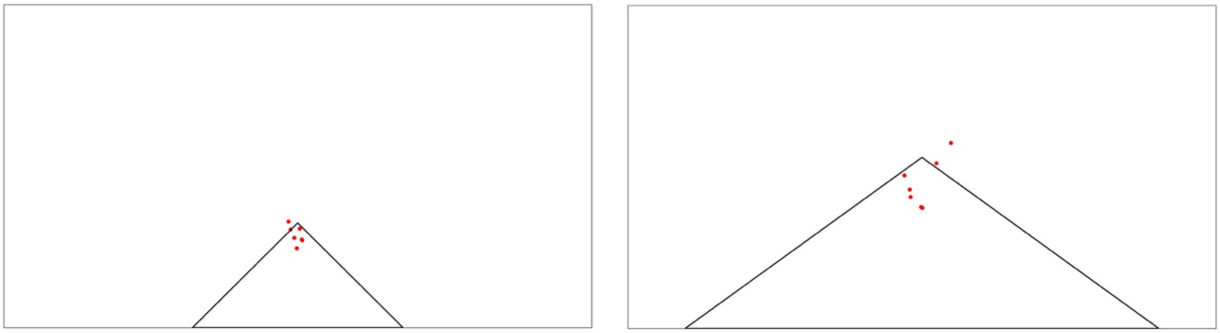


Fig. S2. Localization task responses on a smaller and larger triangle from the two 9-year-old children whose results are also depicted in the middle column of **Fig. S1. Top.** Responses from a child excluded based on the preregistered criterion and **Bottom.** responses from a child included based on the preregistered criterion (also see **Fig. 3**).

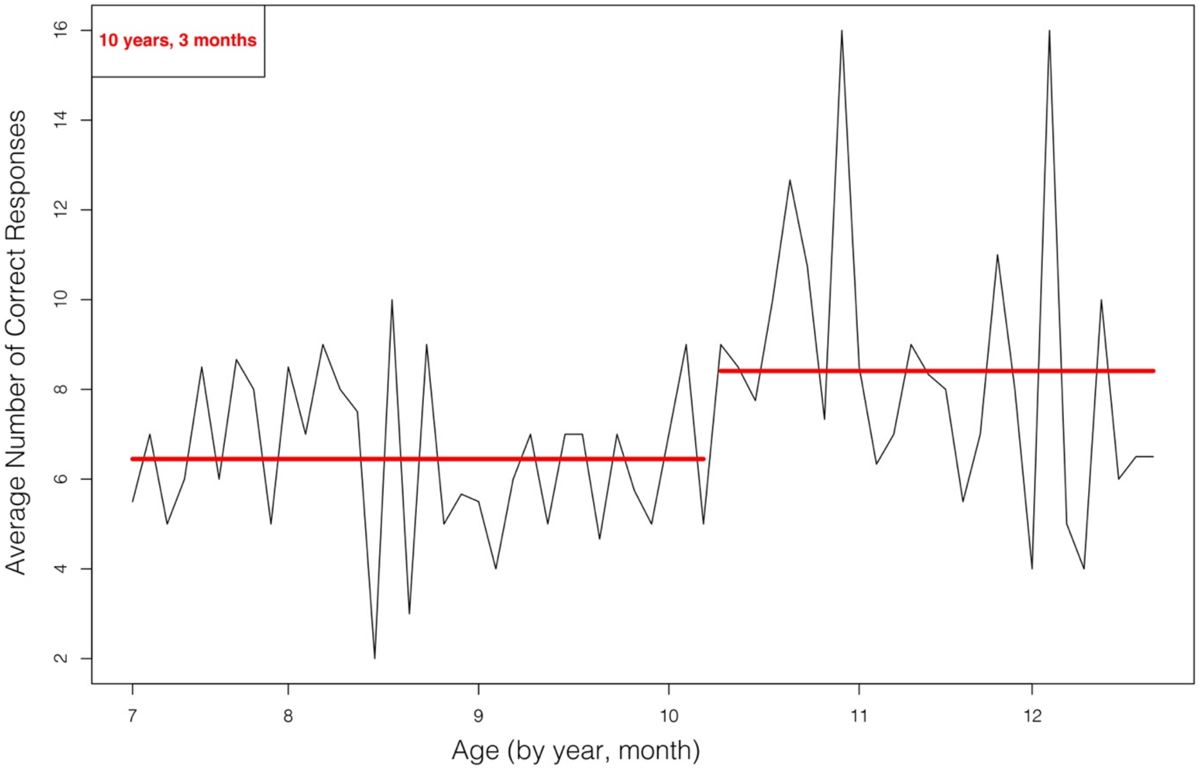


Fig. S3. A change-point analysis on children's accuracy on the reasoning task, with age binned by month and using a binary segmentation method with a Bayesian information criterion (BIC) penalty type, revealed one change point at 10 years 3 months.