AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology


ISSN: 0012-1649

2023, Vol. 59, No. 10, 1784–1793
https://doi.org/10.1037/dev0001600

# Developmental Changes in Drawing Production Under Different Memory Demands in a U.S. and Chinese Sample

Bria Long[1], Ying Wang[2], Stella Christie[2], Michael C. Frank[1], and Judith E. Fan[1, 3]
[1] Department of Psychology, Stanford University
[2] Department of Psychology, Tsinghua University
[3] Department of Psychology, University of California, San Diego

Children's drawings of common object categories become dramatically more recognizable across childhood. What are the major factors that drive developmental changes in children's drawings? To what degree are children's drawings a product of their changing internal category representations versus limited by their visuomotor abilities or their ability to recall the relevant visual information? To explore these questions, we examined the degree to which developmental changes in drawing recognizability vary across different drawing tasks that vary in memory demands (i.e., drawing from verbal vs. picture cues) and with children's shape-tracing abilities across two geographical locations (San Jose, United States, and Beijing, China). We collected digital shape tracings and drawings of common object categories (e.g., cat, airplane) from 4- to 9-year-olds ($N = 253$). The developmental trajectory of drawing recognizability was remarkably similar when children were asked to draw from pictures versus verbal cues and across these two geographical locations. In addition, our Beijing sample produced more recognizable drawings but showed similar tracing abilities to children from San Jose. Overall, this work suggests that the developmental trajectory of children's drawings is remarkably consistent and not easily explainable by changes in visuomotor control or working memory; instead, changes in children's drawings over development may at least partly reflect changes in the internal representations of object categories.

***Public Significance Statement***

The present study examined developmental changes in children's drawings of common object categories from 4 to 9 years of age. The findings suggest relative consistency in the developmental trajectory of drawing recognizability across two different sites (San Jose, United States, and Beijing, China) and two different drawing tasks (verbal vs. picture cues). The digital drawings and stroke-by-stroke records are made available as a resource for future work.

*Keywords:* children's drawings, visual production, tracing, object recognition, visuomotor control



As humans, we have many powerful tools to externalize what we perceive and know, including language and gesture. One tool that has been especially transformative for human cognition and culture is a graphical representation, which allows people to encode their thoughts in a visible, durable format. Drawing is an important case study in graphical representation, being a technique that dates back 60,000 years (Hoffmann et al., 2018), well before the emergence of symbolic writing systems, and is practiced in many cultures.

In modern times, drawings are produced prolifically by children from an early age and change dramatically and systematically across childhood (Karmiloff-Smith, 1990; Kellogg, 1969): younger children (4–5 years of age) tend to include fewer cues in their drawings to differentiate between categories (e.g., child vs. man) than older children (6–7 years of age), who enrich their drawings with more diagnostic part (Sitton & Light, 1992) and relational (Light & Simmons, 1983) information. What are the factors driving these dramatic changes in

Bria Long https://orcid.org/0000-0001-7156-6878
This work was funded by National Science Foundation SPRF-FR Grant 1714726 to Bria Long, a Jacobs Foundation Fellowship to Michael C. Frank, and the Zhou Fund for Language and Cognition. The authors thank Yi Feng and Megan Merrick for their assistance with data collection. The authors also thank the Children's Discovery Museum of San Jose for their collaboration, as well as the parents whose children participated. Thanks to members of the Stanford Language and Cognition Lab for their feedback. The authors declare no conflicts of interest. Data for this research are available at https://osf.io/qxsgw/.

Bria Long served as lead for conceptualization, data curation, formal analysis, visualization, writing–original draft, and writing–review and editing. Ying Wang contributed equally to data curation and resources and served in a supporting role for conceptualization and writing–review and editing. Stella Christie served in a supporting role for data curation, resources, and writing–review and editing. Michael C. Frank served as lead for funding acquisition and served in a supporting role for resources, supervision, and writing–review and editing. Judith E. Fan contributed equally to writing–review and editing and served in a supporting role for formal analysis, software, and writing–original draft. Michael C. Frank and Judith E. Fan contributed equally to conceptualization.

Correspondence concerning this article should be addressed to Bria Long, Department of Psychology, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305, United States. Email: bria@stanford.edu

children's drawings across development? Creating a drawing of even the most familiar concept requires the complex integration of cognitive, motoric, and memory abilities: in order to "draw a [rabbit]," children have to access their mental representation of a "rabbit," maintain this information in memory, and then choose and produce the visual features necessary to convey the identity of this category.

A common view is that these changes are driven primarily by children's increasing ability to plan and control their motor movements (Freeman, 1987; Rehrig & Stromswold, 2018)—rather than any changes in children's internal representations of category information. However, recent work has provided some evidence that changes in children's drawings also reflect changes in children's mental representations of visual concepts. In a large observational data set, older children produced drawings of object categories that were more diagnostic of the categories they were trying to depict (Long et al., 2021). This result held even when accounting for differences in basic shape-tracing abilities and the amount of effort children expended on individual drawings. Furthermore, older children also relied more on these same diagnostic visual features when recognizing other children's drawings. Together, these results suggest that developmental change in these underlying conceptual representations drives parallel changes in both children's production and recognition of drawings. Thus, changes in how well children can produce recognizable drawings might not only reflect improvements in motor control (Akshoomoff & Stiles, 1995; Duval et al., 2015; Vinter & Chartrel, 2008) but perhaps also changes in visual representations (Dekker et al., 2011; Natu et al., 2016).

However, both motoric and visual accounts leave open the possible contribution of children's evolving ability to retrieve and maintain relevant information about categories in working memory. Children's ability to access semantic knowledge about categories and to maintain this information in mind when producing a drawing likely changes across childhood as their working memory capacity increases (Pailian et al., 2016). Here, we directly test the idea that a principal reason younger children produce less recognizable drawings is because they simply have more difficulty recalling the relevant visual features of different categories: that is, when asked to "draw a [rabbit]," they may struggle to conjure up the relevant visual details and then hold in mind what rabbits tend to look like. On this account, providing children with additional visual information about different categories—for example, via canonical photographs of typical exemplars—could help them improve their drawings of these categories, as it can with adults (Fan et al., 2018).

An alternative possibility is that memory constraints are not a major factor that drives developmental changes in children's drawings. On this account, only older children (e.g., 7–9 years old) or adults may be able to produce more recognizable drawings when provided with canonical exemplars from different categories. Supporting this account, prior work also suggests that younger children (e.g., 4–6 years of age) tend to draw what they know about objects rather than integrate information in their immediate perceptual experience (Luquet, 1927). For example, when asked to draw from observation, younger children (e.g., 5- to 6-year-olds) tend to include features that are not visible from their vantage point, yet are diagnostic of category membership (e.g., a handle on a *cup*; Barrett & Light, 1976; Bremner & Moore, 1984), and only omit these features later in development. Similarly, preschoolers will often insist that their nearly identical drawings of different concepts (e.g., *balloon* and *person*) unambiguously refer to different things (Bloom & Markson, 1998).

To tease apart these alternatives, we investigated the development of children's ability to produce recognizable drawings of visual concepts when children were provided with a verbal cue ("can you draw a [rabbit]?") versus when provided with a picture cue ("can you draw this [rabbit] as it looks in the picture?"). On verbal-cue trials, children thus must access their mental representation of a "rabbit" and choose the features necessary to convey that object's identity. Conversely, on picture-cue trials, children are explicitly asked to rely on the visual features provided in a canonical photograph of each object category. If younger children produce more recognizable drawings on the picture versus verbal-cue conditions, this would support an account where younger children's drawings are limited by their ability to recall and maintain information about what visual concepts tend to look like. Conversely, if children's drawings are equally recognizable on picture versus verbal-cue conditions, this would support accounts of developmental changes where memory constraints do not play a major role.

To test the generality of our findings, we recruited children from two sites in different countries—San Jose, United States, and Beijing, China. Most empirical studies on children's drawings have been conducted exclusively on small samples of children from the United States or Western Europe, limiting their generalizability. Furthermore, children in different communities may receive varying degrees of guidance when learning to draw (Huntsinger et al., 2011; Winner, 1989) or acquire different visual conventions when producing drawings of familiar concepts (Cohn, 2012; La Voy et al., 2001; Willats, 2006). Both of these differences may influence the amount and kind of semantic information that drawings generally contain. Here, our aim was to measure developmental changes in the recognizability of children's drawings across these two sites using the same experimental protocol, with the broader goal of laying the groundwork for more comprehensive and controlled investigations of how environmental factors influence the development of drawing production behavior.

In addition, we assessed the degree to which visuomotor development accounts for observed developmental changes in drawing recognizability. While it is uncontroversial that visuomotor control can constrain how and what we can draw, little work has directly related measures of visuomotor control to measures of drawing recognizability (Long et al., 2021). To do so, we measured each child's visuomotor control via a shape-tracing task and related these measurements of tracing accuracy to the recognizability of the drawings that each child produced. As children who learn to write Chinese are also learning to produce complex series of strokes to convey meaning via characters in the Chinese language (McBride, 2016; McBride & Wang, 2015), it is conceivable that visuomotor abilities differ across children within sites, with children in Beijing having more intensive experience producing complex orthographic shapes and showing higher tracing scores (e.g., McBride-Chang et al., 2011), and that any differences in drawing recognizability between sites would be mediated by tracing ability.

In sum, in the current study, we collected shape tracings and digital drawings of visual concepts from 4- to 9-year-old children in Beijing, China and San Jose, United States using both picture cues and verbal cues. In doing so, we make three key contributions to our understanding of the development of children's drawings. First, we replicate prior findings (Long et al., 2021) that the recognizability of children's drawings increases steadily throughout this age range (4–9 years). Second, we test the degree to which memory constraints might account for these developmental changes. In keeping with accounts of intellectual realism and against a large role of limitations on memory, we predicted that only older children would be able to use the visual information present in the canonical photographs to improve their drawings. Third, we test the generalizability of our findings across two sites in different countries. We predicted that we would see convergence in the development of drawing abilities across both geographical sites, with older children becoming progressively better at producing recognizable drawings. In addition, we predicted that most of the variance across geographical sites in drawing ability would be explainable by differences in visuomotor control, operationalized as performance on a shape-tracing task; these primary analyses were preregistered at https://osf.io/3nydc.

## Method

### Participants

Our goal was to recruit a broad sample of 4- to 9-year-old children from our two different sites. Children were recruited from two local children's museums in Northern California (San Jose and Palo Alto) and at preschool and elementary schools in Beijing; approximately equal numbers of participants were recruited in Northern California and the Beijing area. In order to facilitate recruiting in the museum and school contexts, we did not collect other demographic variables other than age from the participants in this study. However, children tested in Beijing were primarily from schools located in middle-class neighborhoods; across prior studies at these same museums in Northern California in our lab in the same year, most children's parents (>50%) reported having "some graduate education" (see Hembacher & Frank, 2020 for detailed statistics from a similar study at these same children's museums).

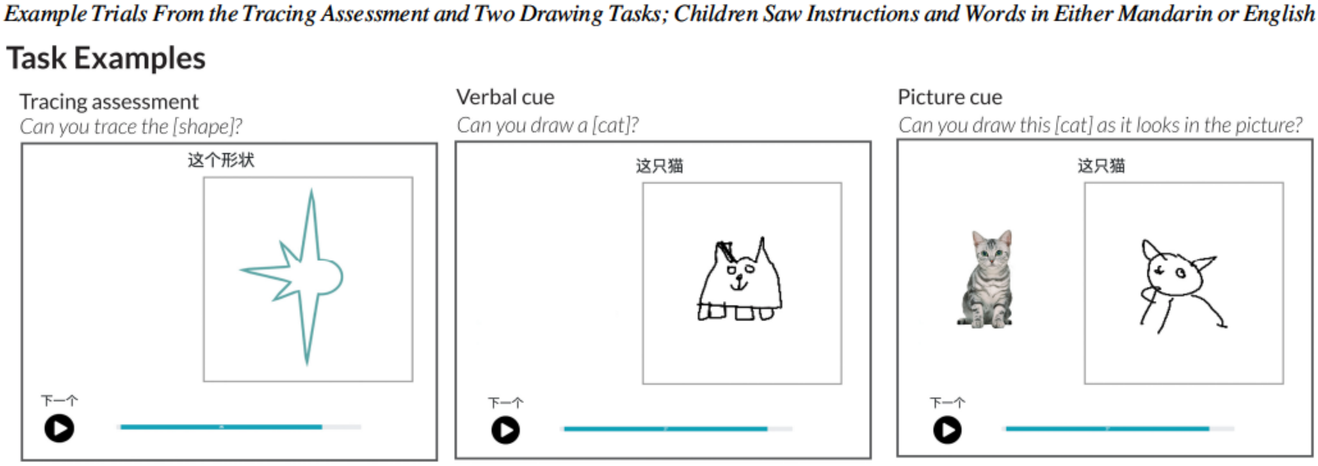
We aimed to recruit 120 children after exclusions (i.e., 20 four-year-olds, 20 five-year-olds, etc.) at each geographical site. In the San Jose sample, 135 children participated in the experiment; six participants were excluded, three for skipping more than six drawing trials and three for scribbling three or more times in a row; these exclusion criteria were specified before beginning data collection (see preregistration). Six additional participants were tested, but their data were not recorded due to a technical error, and two participants never advanced past the practice trials, leading to a final sample of 132 children. In the Beijing sample, 121 children participated; an additional eight participants were tested but their data were not recorded due to a technical error with the remote database. Two children who turned 10 years old between authorization at the school and the date of testing (ages 10 years, 0 months and 10 years, 1 month) were accidentally tested and are thus included in the 9-year-old age group. The number of participants included in each age group at each site after exclusions is shown in Table B1 in Appendix B (range 18–26); Children's age (in years) was included as a continuous numeric predictor in all analyses. On average, each child contributed 11.46 drawings to the analysis (min = 6, max = 12). No additional demographic data were recorded about the participants. This protocol was approved by both the Institutional Review Board at Stanford University (43992, Development of Children's Drawing Abilities) and the Department of Psychology Ethics Committee at Tsinghua University in Beijing, China.

### Stimuli and Task Procedure

At the beginning of the session, a trained experimenter first told each child, "After this game is over, someone is going to try to recognize what you were trying to draw. So, please draw so that someone else could guess what you were trying to draw." A native English speaker gave these instructions to the San Jose sample, and a native Mandarin speaker gave a translation of these instructions to the Beijing sample. Children were then seated in front of a touchscreen tablet (iPad Pro 12.9″) displayed horizontally in a child-safe case at an angle during each drawing session. Children used their fingertips to draw (with no restrictions on which hand/finger), strokes could not be deleted once drawn, and children had to complete each trial within 30 s. Following the initial instructions, children completed two shape-tracing trials. Children first traced a square, then a more complex shape (see shape in Figure 1), providing a baseline measure of visuomotor control in the absence of memory demands. After these tracing trials, children were asked to draw 12 familiar object categories in a random order (see Figure 2; *airplane*, *bike*, *bird*, *car*, *cat*, *chair*, *cup*, *hat*, *house*, *rabbit*, *tree*, and *watch*). Across trials, we manipulated the type of cue (verbal vs. picture) children

**Figure 1**

*Example Trials From the Tracing Assessment and Two Drawing Tasks; Children Saw Instructions and Words in Either Mandarin or English*




*Note.* See the online article for the color version of this figure.

**Figure 2**

*Randomly Sampled, Highly Recognizable Drawings for Each Task, Category, and Geographical Site Made by 6-Year-Old Children*


*Note.* Drawings were randomly sampled. Drawings in Panel A are from the verbal-cue condition, and drawings in Panel B are from the picture-cue condition.

received before producing each drawing (Figure 1), providing a measure of the impact of reminding children of what typical exemplars of each category looked like. Specifically, children first completed six trials with one cue type, then switched to the other cue type. The 12 categories were randomly assigned to each cue type for each participant, and the order in which each cue type was used was counterbalanced across children. On verbal-cue trials, children viewed a short video clip in which an experimenter named the target category: "What about a [cat]? can you draw a [cat]?" On picture-cue trials, children viewed a photograph of a typical exemplar of the target category while listening to an audio clip of the same experimenter who said, "What about this [cat]? can you draw the [cat] as it looks in the picture?" These audio and video stimuli were recorded in Mandarin with a native speaker for the Beijing sample. The photograph then remained on the screen for the duration of the drawing trial. On picture-cued trials, one of three photograph exemplars was randomly sampled for each category for each participant. All experimental code, videos, translations, and stimuli are available on the public repository for this project (Long et al., 2023; https://osf.io/qxsgw/).

## Measuring Drawing Effort Covariates

We recorded both the final drawings and the location of each stroke in a digital format, allowing us to precisely measure several different variables that provided proxies for the amount of effort children invested in producing each drawing. Specifically, we measured (a) the amount of time spent (i.e., end time of last stroke − start time of first stroke), (b) the number of individual strokes drawn (i.e., detected automatically using our web application interface), and (c) the average intensity of each drawing (i.e., proportion of non-white pixels in the drawn image, reflecting the portion of drawing canvas that was filled with "ink").

## Measuring Tracing Accuracy

We used a semiautomated procedure for evaluating how accurately each child performed the tracing task that was validated against empirical judgments of tracing quality, as in prior work (Long et al., 2021), In brief, we decomposed tracing accuracy into two components: a shape error component and a spatial error component. Shape error reflects how closely the participant's tracing matched the contours of the target shape; the spatial error reflects how closely the location, size, and orientation of the participant's tracing matched the target shape. These two error components were computed automatically for each tracing and used to yield a "tracing score" for each tracing that mirrors adult human judgments of tracing quality (see Appendix A for details).

## Measuring Drawing Recognizability

We measured the recognizability of each drawing via an online, 12-Alternative Forced Choice (12-AFC) categorization experiment. Adult participants based in the United States were recruited via Prolific for a 15-min experiment, compensated at $14/hr, and asked to identify the category depicted in a random subset of approximately 140 drawings. Each subset of drawings was balanced with respect to age, category, and site, and each drawing was shown to 10 participants. Participants were shown these drawings in a random sequence and asked, “what does this look like?” Participants selected their responses from the set of 12 categories and were encouraged to provide their best guess if they were unsure. A catch trial was included to verify that participants could accurately describe their goal in this task; no participants were excluded for missing the catch trial. We then computed a recognition score for each drawing, reflecting the proportion of participants who correctly identified the target category.[1]

## Statistical Models

To evaluate our main hypotheses, we fit generalized linear mixed-effects models to the human recognition scores to assess the factors that influenced the recognizability of the drawings that children produced. A first generalized linear mixed effect model was fit to the recognizability scores for each drawing, including fixed effects of children’s age (in years), geographical site (San Jose vs. Beijing), and drawing task (verbal cue vs. picture cue) and the three-way interaction between these key variables. We initially planned to include random slopes for the effect of drawing task on each child (as this varied within-subjects), and random slopes for the effect of the full three-way interactions between task, age, and site on each category. However, models with this random-effects structure failed to converge, and the reported models use the maximal random-effects structure that did converge, which included random slopes for the two-way interaction between task and age on each category. The results of this model are reported in Table 1.

In a secondary analysis, we investigated the degree to which any of the above effects were mediated by children’s tracing abilities or the amount of effort that children expended while drawings. We fit the same model above with additional fixed-effects terms for each child’s estimated tracing score (see Measuring Tracing Abilities), the time that child spent drawing (in seconds), the amount of “ink” used (i.e., percentage of nonwhite pixels in the drawn image), and the number of strokes produced. These additional predictors were first $z$-scored before being included in the model. Finally, we also assessed the degree to which tracing ability development differed across geographical sites, where tracing scores were modeled as a function of age (in years), site, and their interaction, with the same random-effects structure as the first model. The results of this model are reported in Table 2. This analysis plan was preregistered at https://osf.io/3nydc, and data and analysis code are available at https://osf.io/qxsgw (Long et al., 2023).

**Table 1**
*Model Coefficients From a Generalized Linear Mixed Model Predicting the Recognizability of Each Drawing for the Main Experimental Contrasts*

| Fixed effect | Estimate | *SE* | *z* | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 1.99 | 0.16 | 12.17 | <.01 |
| Task | −0.16 | 0.26 | −0.62 | .53 |
| Age | 1.31 | 0.13 | 9.88 | <.01 |
| Site | −0.60 | 0.16 | −3.71 | <.01 |
| Task × Age | −0.08 | 0.19 | −0.41 | .68 |
| Task × Site | 0.19 | 0.16 | 1.17 | .24 |
| Age × Site | −0.15 | 0.16 | −0.91 | .36 |
| Task × Age × Site | −0.06 | 0.16 | −0.38 | .71 |

**Table 2**
*Model Coefficients From a Linear Regression Predicting the Average Tracing Score of Each Child as a Function of Their Age and the Site at Which They Participated*

| Predictor | Estimate | *SE* | *t* | Pr(>\|*t*\|) |
|---|---|---|---|---|
| (Intercept) | 2.73 | 0.07 | 41.30 | .00 |
| Age | 0.46 | 0.07 | 7.01 | <.01 |
| Site | 0.03 | 0.10 | 0.28 | .78 |
| Age × Site | −0.07 | 0.10 | −0.78 | .44 |

# Results

## Confirmatory Analyses

### *Model 1: Main Experimental Contrasts*

We found steady changes in the recognizability of children’s drawings as a function of age (see Figure 3 and Table 1), replicating prior work using the verbal-cue paradigm under less controlled settings (Long et al., 2021). Moreover, we found that this age-related change did not depend on the drawing task children completed: Children’s drawings were just as recognizable when they were cued with a photograph (“draw this [rabbit] as it looks in the picture”) as when they were cued with the category label (“can you draw a [rabbit]?”; see Table 1, no fixed effect of drawing task, $B = -0.16$, $SE = 0.26$, $Z = -0.62$, $p = .53$). While we initially thought that older children might produce more recognizable drawings when cued with a photograph of a highly prototypical exemplar, as adult participants do (Yang & Fan, 2021), we did not find evidence of an interaction between age and cue type (see Table 1, no interaction between fixed effects of age and drawing task; $p = .68$). Unexpectedly, we did observe a fixed effect of geographical site: children in Beijing, China produced drawings that were more recognizable than did children in San Jose, United States (see Table 1, $B = -0.60$, $SE = 0.16$, $Z = -3.71$, $p < .01$).

### *Model 2: Accounting for Tracing Abilities and Effort*

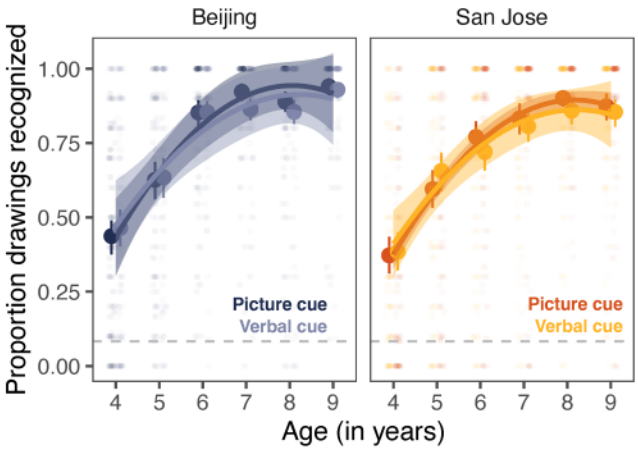
To follow up on this effect of site, we tested two potential sources of variation: visuomotor control and effort. Insofar as differences in visuomotor control could explain differences between sites, we reasoned that this would manifest as a difference in tracing ability across sites that predicts variation in drawing recognizability. Inconsistent with our hypotheses, we found that children across the two sites were indistinguishable with respect to tracing ability on average (no fixed effect of site on tracing scores; see all model coefficients in Table 2; see all individual scores in Figure 4).

We then explored whether there were systematic differences in effort between the two groups, as measured by the amount of time each child spent producing their drawings and the number of strokes drawn. Figure 5 shows three effort covariates—average intensity, number of strokes used, and time spent drawing—measured for each drawing as a function of children’s age, drawing task, and

[1] While in our preregistration we had planned to use automated recognition scores as our main dependent variable, we found that automated recognition scores were only modestly correlated with human recognition ($r = .4$, $p < .001$) and that descriptive plots of the automated recognition scores revealed an overall difference between the two drawing tasks that was not evident in the human recognition data. To be conservative, we thus fit all of our models to the human recognition data.

**Figure 3**

*Proportion of Drawings Recognized as a Function of the Age (in Years) of the Child Who Completed Each Drawing, the Geographical Site They Were Tested at (Beijing vs. San Jose), and the Type of Drawing Task They Completed*

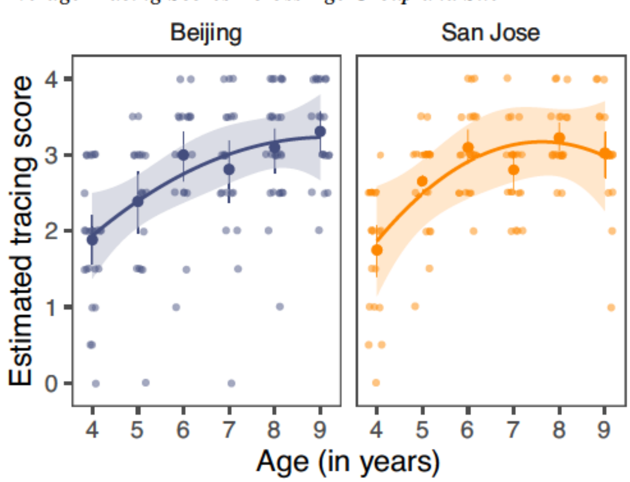



*Note.* Individual data points represent drawings within each condition by an individual participant and are slightly jittered. Error bars show bootstrapped 95% confidence intervals. See the online article for the color version of this figure.

geographical site; these covariates were somewhat correlated with one another (see Table B3 in Appendix B) but not with tracing ability. These exploratory analyses revealed that, overall, children spent more time drawing when prompted with a picture cue than a verbal cue, and that children in the Beijing group spent more time on their drawings than their counterparts in San Jose (see Figure 5; linear mixed effect model predicting drawing duration, fixed effect of site, $B = -0.71$, $SE = 0.07$, $df = 312.67$, $t = -10.02$, $p < .001$, Table B2 in Appendix B). The two other effort covariates

**Figure 4**

*Average Tracing Scores Across Age Group and Site*




*Note.* Each dot represents an average tracing score obtained for each participant and are slightly jittered to show variation. Error bars represent bootstrapped 95% confidence intervals. See the online article for the color version of this figure.

did not show systematic differences that would explain the site differences: Children in San Jose used slightly more ink than children in Beijing but used an equal number of strokes when producing drawings.

If effort and/or tracing ability is sufficient to account for the differences between sites, we reasoned that we should no longer observe a main effect of geographical site on drawing recognizability in an expanded statistical model where we account for these covariates. We thus then included children's average tracing scores and effort covariates measured for each drawing (average intensity, number of strokes used, and time spent drawing) as fixed effects into a second generalized linear mixed-effects model (see Statistical Models). Instead, we still observed a significant effect of geographical site (see Table 3, fixed effect of site, $B = -0.80$, $SE = 0.16$, $Z = -5.04$, $p < .01$), even after having accounted for the effect of individual differences in tracing ability as well as effort covariates for each drawing that was produced. Nonetheless, children's tracing abilities were clearly related to the degree they were able to produce recognizable drawings, the overall correlation between tracing score and average drawing recognizability, $r = .54$, $t(251) = 10.28$, $p < .01$; see also Table 3, fixed effect of tracing ability in expanded model. Together, these results suggest that the site differences in the ability to produce recognizable drawings are not explained by easily measurable differences in effort or tracing ability.

### *Exploratory Analyses*

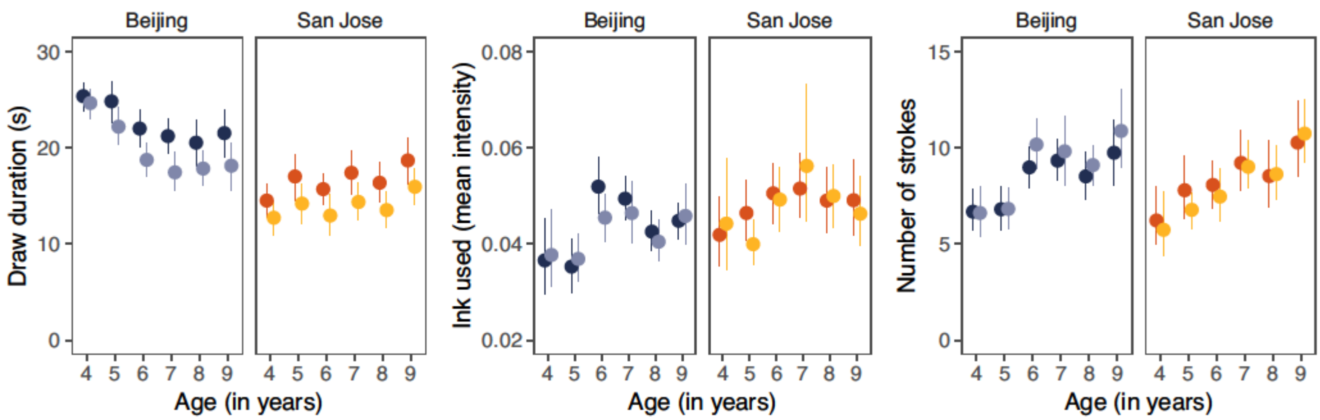
**Item Effects.** In a set of exploratory analyses, we examined how the developmental trajectory for drawing recognizability varied across the 12 object categories we included. For example, some object categories (e.g., cat) may be easier to draw than others (e.g., watch), resulting in shallower or steeper changes in recognizability over age. However, we did not have strong hypotheses about how these trajectories might additionally vary with geographical locations.

Figure 6 shows considerable variation in drawing recognizability across the 12 categories as a function of geographical site and age. For example, certain distinctions between similar categories (e.g., cats and rabbits) appeared more difficult to make, especially for younger children. In this data set, children in the Beijing group appeared to produce more recognizable drawings of certain categories at all ages—including airplanes, birds, and rabbits—while older children in San Jose produced more recognizable drawings of bikes. As a whole, these exploratory analyses provide some converging evidence for systematic differences between object categories and geographical sites. Yet, the interpretation of these item effects is quite complex: children in different locations may spend more or less effort on different categories, which could explain some portion of the measured differences in drawing recognizability. Future work that systematically explores variation across sites may shed light on the degree to which these trends are systematic or dependent on the particular populations tested and photograph exemplars that were used.

However, there are likely to be some differences that are left uncovered by the present set of analyses. Figure 2 shows recognizable, randomly sampled example drawings from 6-year-olds at each category, site, and condition, highlighting potential differences between sites and conditions that may not be captured by a simple 12-AFC recognition metric. For example, some of the drawings made in the picture-cued condition appear more similar to each other (and the picture cue) than drawings made in the verbal-cued condition. Future

**Figure 5**

*Effort Covariates Measured During the Drawing Task—Amount of Time Spent Drawing, Amount of "Ink" Used (i.e., Average Intensity), and Number of Strokes Used—as Function of the Age Group of the Child Who Completed Each Drawing, the Geographical Site They Were Tested at (San Jose vs. Beijing), and the Type of Drawing Task They Completed (Picture Cue vs. Verbal Cue)*




*Note*. Error bars show bootstrapped 95% confidence intervals. See the online article for the color version of this figure.



work that explores hypotheses about how these different tasks and items interact to produce item trajectories will be helpful for understanding what drives variability in children's drawing abilities.

## General Discussion

Here we examined developmental change in children's ability to produce recognizable drawings of familiar visual concepts, and the degree to which this ability is constrained by retrieval of diagnostic attributes from memory. We manipulated memory demands by either cueing children with a category label ("can you draw a [cat]?") or a color photo ("can you draw this [cat] as it looks in the picture?"), and measured the recognizability of each drawing to adult observers. To evaluate the generality of our findings, we recruited two groups of children in different sites: San Jose, United States and Beijing, China.

First, we found steady improvement in the recognizability of children's drawings from 4 to 9 years of age in both groups of children, replicating prior work conducted in a field setting (Long et al., 2021). As in prior work, we also found that these changes in recognizability were not entirely explained by shape-tracing abilities; we still found robust fixed effects of children's age when accounting for individual variation in tracing abilities and effort expended on each drawing. Together with prior work, these findings highlight that children's drawing abilities change gradually throughout middle childhood and raise new questions about the major factors that drive these systematic changes.

Second, inconsistent with our hypotheses, we did not find that the recognizability of children's drawings reliably differed between the two drawing tasks. That is, providing children with reminders of the visual features of each category and asking them to focus on these details on picture-cued trials did not lead children to produce drawings that were more recognizable. Thus, these results suggest that younger children's drawings of familiar visual concepts are unlikely to be constrained by their ability to recall category-diagnostic visual features when they are trying to draw them. Future work that systematically manipulates the kinds of visual information that children are provided with (e.g., photographs vs. iconic drawings) and the instructions that children receive will confirm the generality of this finding.

We also discovered that drawing recognizability differed between the geographical sites: children in the Beijing group produced more recognizable drawings than those in the San Jose group. These intriguing group-level differences are consistent with a wide variety of potential explanations. One possibility is that because children in the Beijing group drew at their preschool and elementary schools, whereas children in the San Jose group drew in a room at children's science museums, these two testing environments could have induced motivational differences that affected how much effort children put into their drawings. We did not collect additional demographic information from our participants, and one of several possibilities is that children in Beijing had different socioeconomic backgrounds that led to higher motivation to produce accurate drawings. Indeed, children in Beijing took more time on average to complete their drawings. However, contrary to a strong version of this account, we found that

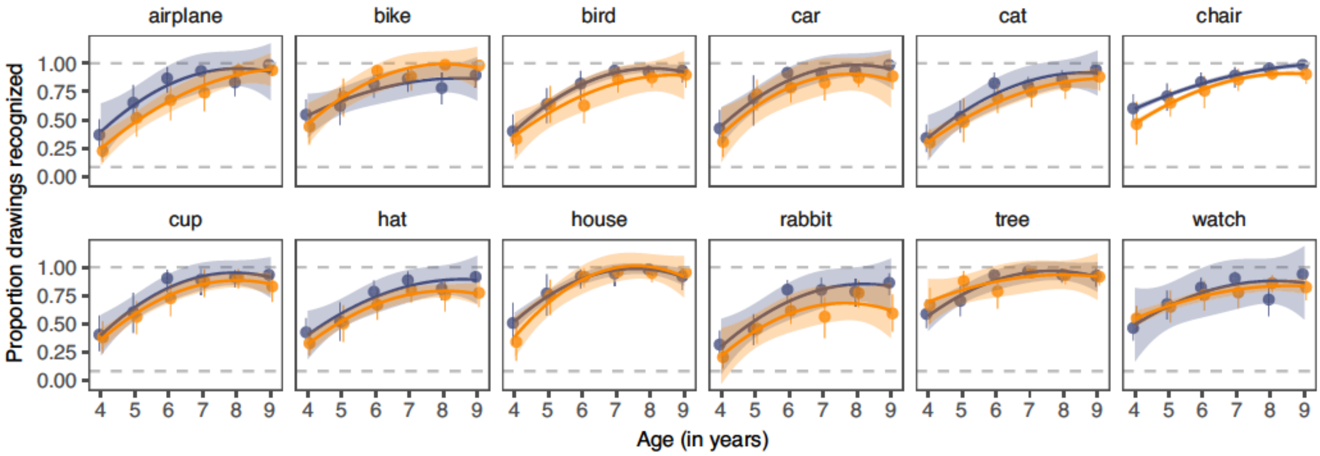
**Table 3**

*Model Coefficients From a Generalized Linear Mixed Model Predicting the Recognizability of Each Drawing as a Function the Both the Main Experimental Contrasts (Task, Site, and Age) as Well as Several Effort Covariates and Estimates of Children's Tracing Abilities*

| Fixed effect | Estimate | *SE* | *z* | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 2.14 | 0.16 | 13.72 | <.01 |
| Task | −0.26 | 0.26 | −0.98 | .33 |
| Age | 1.01 | 0.14 | 7.31 | <.01 |
| Site | −0.80 | 0.16 | −5.04 | <.01 |
| Estimated tracing score | 0.35 | 0.07 | 4.82 | <.01 |
| Average intensity | 0.06 | 0.02 | 2.47 | .01 |
| Draw duration | −0.25 | 0.03 | −7.58 | <.01 |
| Number of strokes | 0.33 | 0.03 | 10.99 | <.01 |
| Task × Age | −0.10 | 0.19 | −0.54 | .59 |
| Task × Site | 0.22 | 0.16 | 1.38 | .17 |
| Age × Site | −0.04 | 0.16 | −0.22 | .82 |
| Task × Age × Site | −0.07 | 0.16 | −0.43 | .67 |

**Figure 6**

*Developmental Trajectory of Drawing Recognizability for Each Category and for Each Geographical Site*




*Note.* Error bars represent 95% bootstrapped confidence intervals. See the online article for the color version of this figure.

the substantial individual variation in basic shape-tracing ability and effort (as assessed via average intensity, time, and strokes spent while drawing) was insufficient to account for differences in drawing recognizability between groups.

What additional factors may explain the difference in drawing recognizability across sites? A second possibility is that the children who are learning to produce the complex, visually demanding characters in the Chinese language that have meaningful subparts (McBride, 2016; McBride & Wang, 2015) may develop more sophisticated visuomotor skills that are recruited while drawing. For example, the skills necessary to generate, combine, and arrange multiple shape parts in specific locations may be well practiced by children learning to produce Chinese characters, and not well captured by the relatively simple shape-tracing assessment that we used. Future work that measures the development of writing, tracing, and drawing abilities within individual children learning to write in different languages will be useful for understanding the degree to which children's writing and drawing abilities scaffold each other more generally.

Finally, we also found that the groups differed with respect to which visual concepts were easier for children to convey in their drawings, perhaps reflecting different patterns of exposure to these concepts or different amounts of practice drawing these visual concepts. Indeed, another natural question for future work concerns the relationship between the kind of experience children have with these visual concepts and the way they produce drawings of them. For example, if some children are exposed to a wide variety of illustrations of some concepts but not others (e.g., in books or other media), they may also be able to produce more recognizable drawings of those concepts. Conversely, if some children spend more time drawing certain visual concepts but not others (e.g., in school or at home), they may also be better at recognizing graphical representations of them in other contexts, and perhaps explicitly identifying their diagnostic features. Prior work has found that older children not only produce more recognizable drawings of object categories, but also tend to be better at recognizing other children's drawings (Long et al., 2021), providing evidence that visual production and recognition abilities are related throughout childhood at the group level. However, further investigation will be necessary for understanding what kinds of experience are responsible for concurrent developmental changes in these two behaviors at the individual level.

Thus, further systematic measurement of drawing behavior across a wider variety of geographical, socioeconomic, and cultural contexts will be crucial for producing more robust and precise estimates of developmental variability, critical to strongly evaluate causal theories of such variation (Amir & McAuliffe, 2020; Cao et al., 2022; Frank et al., 2021). Without such data sets, it can be tempting to draw unwarranted causal inferences about the impact of single demographic covariates, such as nationality (Kuwabara & Smith, 2016; Winner, 1989). We have thus made our drawing data set publicly available to contribute to this more cumulative effort and also to reduce barriers to investigation of other aspects of these drawings other than their recognizability, such as which visual features children prioritized in their drawings, the order in which they drew them, as well as variation in visual style (Gernhardt et al., 2015; Senzaki et al., 2014).

Overall, our results thus suggest that children's ability to convey their knowledge about familiar visual concepts changes systematically with age across two different geographical contexts and is unlikely to be constrained by limitations in children's ability to retrieve the relevant information associated with these concepts and maintain this information in working memory. We propose that further investigation into the factors that influence how children learn to identify and combine the relevant diagnostic features of different visual concepts will further explain individual and age-related variation in drawing abilities. More broadly, we believe that this approach to concurrently quantifying multiple sources of variation in the context of rich, naturalistic behaviors will lead to more robust and unified theories of cognitive development.

## References

Akshoomoff, N. A., & Stiles, J. (1995). Developmental trends in visuospatial analysis and planning: I. Copying a complex figure. *Neuropsychology*, *9*(3), 364–377. https://doi.org/10.1037/0894-4105.9.3.364

Amir, D., & McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior*, *41*(5), 430–444. https://doi.org/10.1016/j.evolhumbehav.2020.06.006

Barrett, M., & Light, P. (1976). Symbolism and intellectual realism in children's drawings. *British Journal of Educational Psychology*, *46*(2), 198–202. https://doi.org/10.1111/j.2044-8279.1976.tb02312.x

Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science*, *9*(3), 200–204. https://doi.org/10.1111/1467-9280.00038

Bremner, J. G., & Moore, S. (1984). Prior visual inspection and object naming: Two factors that enhance hidden feature inclusion in young children's drawings. *British Journal of Developmental Psychology*, *2*(4), 371–376. https://doi.org/10.1111/j.2044-835X.1984.tb00944.x

Cao, A., Carstensen, A., Gao, S., & Frank, M. C. (2022, December 22). *US-China differences in cognition and perception across 12 tasks: Replicability, robustness, and within-culture variation*. https://doi.org/10.31234/osf.io/bxqj3.

Cohn, N. (2012). Explaining "I can't draw": Parallels between the structure and development of language and drawing. *Human Development*, *55*(4), 167–192. https://doi.org/10.1159/000341842

Dekker, T., Mareschal, D., Sereno, M. I., & Johnson, M. H. (2011). Dorsal and ventral stream activation and object recognition performance in school-age children. *NeuroImage*, *57*(3), 659–670. https://doi.org/10.1016/j.neuroimage.2010.11.005

Duval, T., Rémi, C., Plamondon, R., Vaillant, J., & O'Reilly, C. (2015). Combining sigma-lognormal modeling and classical features for analyzing graphomotor performances in kindergarten children. *Human Movement Science*, *43*, 183–200. https://doi.org/10.1016/j.humov.2015.04.005

Fan, J. E., Yamins, D. L., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive Science*, *42*(8), 2670–2698. https://doi.org/10.1111/cogs.12676

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The wordbank project*. MIT Press. https://doi.org/10.7551/mitpress/11577.001.0001

Freeman, N. H. (1987). Current problems in the development of representational picture-production. *Archives de Psychologie*, *55*(213), 127–152.

Gernhardt, A., Rübeling, H., & Keller, H. (2015). Cultural perspectives on children's tadpole drawings: At the interface between representation and production. *Frontiers in Psychology*, *6*, Article 812. https://doi.org/10.3389/fpsyg.2015.00812

Hembacher, E., & Frank, M. C. (2020). The early parenting attitudes questionnaire: Measuring intuitive theories of parenting and child development. *Collabra: Psychology*, *6*(1), 16. https://doi.org/10.1525/collabra.190

Hoffmann, D. L., Standish, C. D., García-Diez, M., Pettitt, P. B., Milton, J. A., Zilhão, J., Alcolea-González, J. J., Cantalejo-Duarte, P., Collado, H., De Balbín, R., & Lorblanchet, M. (2018). U-Th dating of carbonate crusts reveals Neanderthal origin of Iberian cave art. *Science*, *359*(6378), 912–915. https://doi.org/10.1126/science.aap7778

Huntsinger, C. S., Jose, P. E., Krieg, D. B., & Luo, Z. (2011). Cultural differences in Chinese American and European American children's drawing skills over time. *Early Childhood Research Quarterly*, *26*(1), 134–145. https://doi.org/10.1016/j.ecresq.2010.04.002

Karmiloff-Smith, A. (1990). Constraints on representational change: Evidence from children's drawing. *Cognition*, *34*(1), 57–83. https://doi.org/10.1016/0010-0277(90)90031-E

Kellogg, R. (1969). *Analyzing children's art*. National Press Books.

Kuwabara, M., & Smith, L. B. (2016). Cultural differences in visual object recognition in 3-year-old children. *Journal of Experimental Child Psychology*, *147*, 22–38. https://doi.org/10.1016/j.jecp.2016.02.006

La Voy, S. K., Pedersen, W. C., Reitz, J. M., Brauch, A. A., Luxenberg, T. M., & Nofsinger, C. C. (2001). Children's drawings: A cross-cultural analysis from Japan and the United States. *School Psychology International*, *22*(1), 53–63. https://doi.org/10.1177/0143034301221005

Light, P., & Simmons, B. (1983). The effects of a communication task upon the representation of depth relationships in young children's drawings. *Journal of Experimental Child Psychology*, *35*(1), 81–92. https://doi.org/10.1016/0022-0965(83)90071-1

Long, B., Fan, J., Chai, Z., & Frank, M. C. (2021). *Parallel developmental changes in children's drawing and recognition of visual concepts*. PsyArXiv. https://doi.org/10.31234/osf.io/5yv7x

Long, B., Wang, Y., Christie, S., Frank, M. C., & Fan, J. F. (2023). *Developmental changes in drawing production under different memory demands in a U.S. and Chinese sample*. https://osf.io/qxsgw/

Luquet, G. H. (1927). *Le dessin enfantin (Bibliothèque de psychologie de l"enfant et de pédagogie)* [Children's drawing]. Alcan.

McBride, C. (2016). Is Chinese special? Four aspects of Chinese literacy acquisition that might distinguish learning Chinese from learning alphabetic orthographies. *Educational Psychology Review*, *28*(3), 523–549. https://doi.org/10.1007/s10648-015-9318-2

McBride, C., & Wang, Y. (2015). Learning to read Chinese: Universal and unique cognitive cores. *Child Development Perspectives*, *9*(3), 196–200. https://doi.org/10.1111/cdep.12132

McBride-Chang, C., Zhou, Y., Cho, J. R., Aram, D., Levin, I., & Tolchinsky, L. (2011). Visual spatial skill: A consequence of learning to read? *Journal of Experimental Child Psychology*, *109*(2), 256–262. https://doi.org/10.1016/j.jecp.2010.12.003

Natu, V. S., Barnett, M. A., Hartley, J., Gomez, J., Stigliani, A., & Grill-Spector, K. (2016). Development of neural sensitivity to face identity correlates with perceptual discriminability. *The Journal of Neuroscience*, *36*(42), 10893–10907. https://doi.org/10.1523/JNEUROSCI.1886-16.2016

Pailian, H., Libertus, M. E., Feigenson, L., & Halberda, J. (2016). Visual working memory capacity increases between ages 3 and 8 years, controlling for gains in attention, perception, and executive control. *Attention, Perception, & Psychophysics*, *78*(6), 1556–1573. https://doi.org/10.3758/s13414-016-1140-5

Rehrig, G., & Stromswold, K. (2018). What does the DAP:IQ measure? Drawing comparisons between drawing performance and developmental assessments. *The Journal of Genetic Psychology*, *179*(1), 9–18. https://doi.org/10.1080/00221325.2017.1392281

Sandkühler, R., Jud, C., Andermatt, S., & Cattin, P. C. (2018). *AirLab: autograd image registration laboratory*. arXiv preprint arXiv:1806.09907.

Senzaki, S., Masuda, T., & Nand, K. (2014). Holistic versus analytic expressions in artworks: Cross-cultural differences and similarities in drawings and collages by Canadian and Japanese school-age children. *Journal of Cross-Cultural Psychology*, *45*(8), 1297–1316. https://doi.org/10.1177/0022022114537704

Sitton, R., & Light, P. (1992). Drawing to differentiate: Flexibility in young children's human figure drawings. *British Journal of Developmental Psychology*, *10*(1), 25–33. https://doi.org/10.1111/j.2044-835X.1992.tb00560.x

Vinter, A., & Chartrel, E. (2008). Visual and proprioceptive recognition of cursive letters in young children. *Acta Psychologica*, *129*(1), 147–156. https://doi.org/10.1016/j.actpsy.2008.05.007

Willats, J. (2006). *Making sense of children's drawings*. Psychology Press.

Winner, E. (1989). How can Chinese children draw so well? *Journal of Aesthetic Education*, *23*(1), 41–63. https://doi.org/10.2307/3332888

Yang, J., & Fan, J. E. (2021). *Visual communication of object concepts at different levels of abstraction*. arXiv Preprint arXiv:2106.02775. https://doi.org/10.48550/arXiv.2106.02775

(*Appendices follow*)

# Appendix A

## Tracing Score Calculation

To compute the tracing error components, we applied an image registration algorithm, AirLab (Sandkühler et al., 2018), to align each tracing to the target shape, yielding an affine transformation matrix that minimized the pixel-wise correlation distance between the aligned tracing, *T*, and the target shape,

$$S : \text{Loss}_{\text{NCC}} = -\frac{\sum S \cdot T - \sum E(S)E(T)}{N \sum \text{Var}(S)\text{Var}(T)}, \quad \text{(A1)}$$

where *N* is the number of pixels in both images. The shape error was defined by the final correlation distance between the aligned tracing and the target shape. The spatial error was defined by the magnitude of three distinct error terms: location, orientation, and size error, derived by decomposing the affine transformation matrix above into translation, rotation, and scaling components, respectively. In sum, this procedure yielded four error values for each tracing: one value representing the shape error (i.e., the pixel-wise correlation distance) and three values representing the spatial error (i.e., magnitude of translation, rotation, scaling components).

We used the tracing quality ratings obtained in Long et al. (2021) to assign weights to each of their error terms; adult observers ($N = 70$) rated 1,325 tracings (i.e., 50–80 tracings per shape per age) and evaluated "how well the tracing matches the target shape and is aligned to the position of the target shape" on a 5-point scale. An ordinal regression mixed-effects model to predict these 5-point ratings, which contained correlation distance, translation, rotation, scaling, and shape identity (square vs. star) as predictors, with random intercepts for rater. This model yielded parameter estimates that could then be used to score each tracing in the data set; we averaged scores for both shapes to yield a single tracing score for each participant.

# Appendix B

## Additional Analyses

**Table B1**

*Number of Participants Included in the Final Data Set Within Each Combination of Site and Age Group*

| Age group | Site | Number of participants |
|---|---|---|
| 4-year-olds | Beijing | 26 |
| 4-year-olds | San Jose | 20 |
| 5-year-olds | Beijing | 18 |
| 5-year-olds | San Jose | 20 |
| 6-year-olds | Beijing | 20 |
| 6-year-olds | San Jose | 21 |
| 7-year-olds | Beijing | 21 |
| 7-year-olds | San Jose | 20 |
| 8-year-olds | Beijing | 26 |
| 8-year-olds | San Jose | 20 |
| 9-year-olds | Beijing | 21 |
| 9-year-olds | San Jose | 20 |

**Table B2**

*Model Coefficients From a Linear Mixed Model Predicting How Long Children Spent on Each Drawing as a Function of Site (Beijing vs. San Jose) and Drawing Condition*

| Predictor | Estimate | *SE* | *df* | *t* | Pr(>\|*t*\|) |
|---|---|---|---|---|---|
| (Intercept) | 0.64 | 0.13 | 20.95 | 4.98 | .00 |
| Site | −0.71 | 0.07 | 312.67 | −10.02 | <.01 |
| Condition | −0.31 | 0.03 | 2,640.23 | −9.28 | <.01 |
| Site × Condition | 0.00 | 0.05 | 2,636.34 | −0.09 | .93 |

**Table B3**

*Pearson's Correlation Coefficients for All Effort Covariates (Estimated for Each Drawing) and Tracing Scores (Estimated by Each Participant)*

| | Number of strokes | Drawing duration (s) | Mean intensity | Average tracing score |
|---|---|---|---|---|
| Number of strokes | — | .51 | .50 | .09 |
| Drawing duration (s) | .51 | — | .27 | −.03 |
| Mean intensity | .50 | .27 | — | −.01 |
| Average tracing score | .09 | −.03 | −.01 | — |

Received August 24, 2022
Revision received June 15, 2023
Accepted June 21, 2023 ■