Consistency and variability in the development of children's drawings between two countries

Bria Long[1], Ying Wang[2], Stella Christie[2], Michael C. Frank[1], & Judith E. Fan[3]

[1] Stanford University

[2] Tsinghua University

[3] University of California, San Diego

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Bria Long, 450 Jane Stanford Way, Stanford CA 94305. E-mail: bria@stanford.edu

Abstract

Childen's drawings of common object categories become dramatically more recognziable across development. What are the major factors that explain these developmental changes? Here, we examined the degree to which these changes in recognzibility vary across different drawing tasks (i.e. drawing from observation vs. from memory), geographical locations, and children's tracing abilities. To do so, we collecting digital drawings of object categories (e.g., cat, airplane) from 4-9 year-olds (N=253) who were recruited near San Jose, USA and Beijing, China. Overall, we show broad consistency in these developmental trajectories across object categories, drawing tasks, and these two geographical locations, and we find that children's tracings abilites are good predictors of the recognizability of the drawings that they produce. In addition, we find that children recruited near Beijing, China produced overal more recognizable drawings but showed similar tracing abilities to childen recruited near San Jose, CA. Overall, this work suggests that the developmental trajectory of children's drawings are remarkably consistent and not easily explainable by changes in their visuomotor control.

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

Word count: X

Consistency and variability in the development of children's drawings between two countries

As humans, we have many powerful tools to externalize what we know, including language and gesture. One tool that has been transformative for human cognition and culture is 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. Figurative drawings have long provided inspiration for scientists investigating children's emerging cognitive abilities (Minsky & Papert, 1972), and accordingly a long history of work has examined changes in children's drawings across development (Fury, Carlson, & Sroufe, 1997; Karmiloff-Smith, 1990; Kellogg, 1969; Piaget, 1929). Indeed, there appear to be dramatic changes in how children encode diagnostic visual information in their drawings across age; younger children (4-5 years) tend to include fewer cues in their drawings to differentiate between target concepts (e.g., *adult* vs. *child*) than older children, who enrich their drawings with more diagnostic part (Sitton & Light, 1992) and relational (Light & Simmons, 1983) information. Overall, younger children produce drawings of object categories that are less diagnostic of the categories they are trying to depict (Long, Fan, & Frank, 2018).

What drives these dramatic changes in children's drawings across development? 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). While such changes in visuomotor control are clearly important, this view fails to account for other important constraints, such as how well children are able to access previously acquired semantic knowledge about each target concept and maintain this information in working

memory during drawing production. % how recognizable drawing is a complex skill that integrates many different aspects of cognition. % For example, in order to "draw a [rabbit]", a child has to access their mental representation of a rabbit, choose the particular visual features needed to convey this visual concept, and then maintain this representation and goal in mind as they plan and produce a series of motor movements. Thus, drawing is a skill that likely also relies on children's evolving perceptual category representations (Dekker, Mareschal, Sereno, & Johnson, 2011; Long et al., 2018; Natu et al., 2016) as well as their increasing working memory capacity (Pailian, Libertus, Feigenson, & Halberda, 2016).

One idea is that a principal reason younger children produce less recognizable drawings is because they simply have more difficulty recalling the relevant perceptual features of different categories: that is, when asked to "draw a [rabbit]", they may struggle to conjure up the relevant perceptual details and then hold in mind what rabbits tend to look like. On this account, providing children with additional perceptual information about different categories – for example, canonical photographs of typical exemplars – could help them improve their drawings of these categories, as it does with adults (Fan et al., 2020). However, prior work also suggests that younger children tend to draw what they know about objects rather than integrate the information in their immediate perceptual experience.

For example, when asked to draw from observation, younger children tend to include features that are not visible from their vantage point, yet are diagnostic of category membership (e.g., a handle on a *mug*) (Barrett & Light, 1976; Bremner & Moore, 1984), and only omit these features later in development. Similarly, young children will insist that their nearly identical drawings of different concepts (e.g., balloon and person) unambiguously refer to different things (Bloom & Markson, 1998). Thus, an alternative possibility is that only older children may be able to produce more recognizable drawings when provided with canonical exemplars of different categories. On this account, changes in children's drawings from memory across age may be largely due to other factors beyond changes in memory – for example, changes in how children represent the diagnostic visual features of each category

(Long et al., 2018).

To tease apart these alternatives, we investigated the development of children's ability to produce recognizable drawings of visual concepts when asked to draw from memory ("Can you draw a [rabbit]?) and when asked to draw a canonical exemplar from observation ("Can you draw this [rabbit] as it looks in the picture?") across two different geographical sites (San Jose, USA and Beijing, China). To do so, we gathered digital drawings from a large sample of children 4-9 years of age in a controlled experimental setting. While children in different communities may spend more or less time practicing drawing or use different visual conventions to draw (Cohn, 2012; Willats, 2006), most empirical studies on children's drawings have been conducted exclusively on children from the United States – or focused on differences in educational practices between communities (Huntsinger, Jose, Krieg, & Luo, 2011; La Voy et al., 2001; Winner, 1989). To assess the degree to which visuomotor development accounts for any observed variation, we also measured each child's shape tracing abilities.

We had several predictions about the nature of the effects that we would observe. First, we predicted that only older children would be able to use the visual information present in the canonical photographs to improve their drawings, in keeping with accounts of naïve realism and contra a strong account of working memory limitations. Second, 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. Finally, 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 pre-registered at https://osf.io/qymjr/.

Overall, we replicated prior work, showing strong and consistent developmental changes in the recognizability of children's drawings across development.

Methods

Participants

265 participants were recruited from the San Jose Children's Discovery museum, the Palo Alto Junior Museum and Zoo, and preschool and elementary schools outside of Beijing; approximately equal numbers of participants were recruited in Northern California and the Beijing area. We aimed to recruit approximately 120 children between 4-9 years of age after exclusions (i.e. 20 4-year-olds, 20 5-year-olds, etc.). In the US-based sample, 135 children participated; 6 participants were excluded, (3) for skipping more than 6 trials, and (3) for scribbling three or more times in a row; (6) participants were tested but their data was not recorded due to a technical error, and (2) participants never made it past the practice trials, leading to a final sample of 121 children. In the China based sample, 132 children participated; an additional 8 participants were tested but their data was not recorded due to a technical error with the remote database. Two 10-year-olds (aged 10 years, 0 months and 10 years, 1 month) were accidentally tested and included in the 9-year-old age group. A complete breakdown of the number of subjects in each age group and site can be found in the Appendix, Table 1A.

Blank drawings were excluded from analysis, and drawings were randomly undersampled to have an equal number of drawings for each leave-one-out-classification. On average, each child contributed 11.46 to analysis (min 6, max = 12).

Task Procedure

Before beginning, 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 try what you were trying to draw." Children then completed a series of

tracing trials designed to both familiarize them with the tablet interface and to obtain an estimate of their tracing abilities; children were first asked to trace a square and then a complex shape (see Figure X). Children used their fingers to draw and had a maximum of 30 seconds to produce their tracings/drawings on all trials; strokes could not be deleted once drawn.

The primary experimental manipulation was the type of cue used to prompt children to produce a drawing. Specifically, there were two cue types: (1) a verbal cue provided by the experimenter in a 5s video clip (2) an photo cue which remained on screen continuously while children drew. In the verbal-cue condition, children heard "What about a [x]? Can you draw a [x]?" In the image-cue condition, children heard audio of the same experimenter said, "What about this [x]? Can you draw the [x] as it looks in the picture?". Children completed 12 drawing trials, with six categories each randomly assigned to each condition and displayed in a random order. For photo-cues trials, one of three possible exemplars was randomly chosen (see Appendix, Figure X for all stimuli). Children completed drawing trials were blocked by these two conditions; the condition order counterbalanced across subjects in each age group. All instructions were translated into Mandarin by a native speaker; see Appendix, Table X for exact translations.

Measuring Tracing Accuracy

As in (Long et al., 2018), we used an automated procedure for evaluating how accurately participants performed the tracing task, validated against empirical judgments of tracing quality. 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.

To compute these error components, we applied an image registration algorithm, AirLab (Sandkühler, Jud, Andermatt, & Cattin, 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:

$Loss_{NCC} = -\frac{\sum S \cdot T - \sum E(S) E(T)}{N \sum Var(S) Var(T)}$, 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 same tracing quality ratings to obtained in (Long et al., 2018) to assign weights to each of their error terms; adult observers (N=70) rated 1325 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 dataset; we averaged scores for both shapes within session to yield a single tracing score for each participant.

Measuring effort covariates

For each drawing trial, children had up to 30 seconds to complete their drawings with their fingers. We recorded both the final drawings and the parameters of each stroke produced by children, allowing us to estimate the amount of time children put into their

drawings (e.g., end time of last stroke — start time of first stroke). As a second measure of effort, we also counted the number of strokes that children put into a given drawing. Finally, we estimated the proportion of the drawing canvas that was filled (e.g., "ink used") by computing the proportion of each final drawing that were non-white pixels.

Measuring drawing recognizability

Human recognition scores. We assessed the recognizability of each drawing via a recognition experiment. Adult participants were recruited via Prolific for a 15 minute experiment and compensated at $14/hour. Participants were shown a random subset of the drawings and asked "What does this look like?" and then selected an answer from the twelve categories that children were prompted to draw. Participants were encouraged to guess.

Automated recognition scores. We used a combination of neural neural activations and logistic regressions to obtain automated recognition scores, as per our pre-registered protocol. To encode the high-level visual features of each sketch, we used the VGG-19 architecture [**?**], a deep convolutional neural network pre-trained on Imagenet classification. We used model activations in the second-to-last layer of this network, which is the first fully connected layer of the network (FC6). Raw feature representations in this layer consist of flat 4096-dimensional vectors, to which we applied channel-wise normalization across all filtered drawings in the dataset. Next, we used these features to train object category decoders. We then trained a 12-way logistic classifier with L2 regularization (tolerance = .1, regularization = .1), and used this classifier to estimate the category label for each drawing in the dataset.To avoid any bias due to imbalance in the distribution of drawings over categories, we randomly under sampled such that there were an equal number of drawings for each combination of geographical site (San Jose, Beijing) and the 12 categories. No additional metadata about the age of the child who produced each sketch was provided to the decoder. This procedure was repeated for each drawing in the dataset,

yielding both a binary a recognition score and the softmax probability assigned to each target class in the dataset.

Statistical models

Two mixed effects models were fit to assess the factors that influenced the recognizability of the drawings that children produced. A first linear mixed effect model was fit to the recognizability scores assigned to each drawing, including fixed effects of children's age (in years), geographical site (San Jose vs. Beijing), and drawing task (photo-cue vs. label-cue), and the three way interaction between these key variables. We initially attempted to include random slopes for the effect of drawing task on each child (as this varied within-subjects), and random slopes for the effect the full three-way interactions between task, age, and site on each category. However, all models with this complicated random effects structure failed to converge, and the reported models the maximal random effects structure that did converge following (**???**). We then ran a second model predicting the human recognition scores for each drawing with the same fixed and random effect structure.

In a secondary analysis, we aimed to understand the degree to which any of the above effects were mediated by (1) children's tracing abilities and (2) the amount of effort that children expended while drawings. We thus ran the same main model while also now including fixed effects of children's estimated tracing score (see above), the time children spent drawing (in seconds), the mean intensity of the drawing (i.e. percentage of non-white pixels), and the number of strokes children used. All predictors were scaled to have a mean of 0 and a standard deviation of 1. Finally, we also assessed the degree to which tracing ability development differed by geographical site, where tracing scores were modeled as a function of age (in years), site, and their interaction, with random intercepts for each participant.

Results

Overall, we found relative consistency in the developmental trajectory of children's drawings. Figure **??** shows the recognizability of children's drawings at each age as a function of the drawing task and the geographical site that they were located in. We found no difference in the recognizability of children's drawings according to the task they were asked to do: drawings were equally recognizable when they were asked to "Draw this rabbit as it looks in the picture" and they were simply asked "What about a rabbit? Can you draw a rabbit." While we anticipated that older children might produce more recognizable drawings when drawing from observation – as do some adults when presented with canonical photographs of object categories (**???**) – this was not the case. All model coefficients can be found in Table 1.

However, we did observe a main effect of geographical site: Children who were recruited outside of Beijing, China tended to produce more recognizable drawings than children who were recruited in San Jose, USA. In a second set of analyses, we aimed to understand the nature of this difference. We hypothesized that any differences in the recognizability of children's drawings across sites would be mediated by the amount of effort children expended while drawing or their tracing abilities. To test this hypothesis, we thus 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 the same linear mixed-effects model. If these effort covariates explained the differences between sites that we observed, we reasoned that we should no longer observe a main effect of geographical location on drawing recognizability. Contra this hypothesis, however, we still observed a significant effect of geographical site (see Table 2), despite the fact that individual children's tracing abilities were strong predictors of how recognizable their drawings were.

```
## % latex table generated in R 3.6.1 by xtable 1.8-4 package
```

```
## % Wed Jun 23 21:21:29 2021

## \begin{table}[ht]

## \centering

## \begin{tabular}{rrrrrr}

##    \hline

##  & Estimate & Std. Error & df & t value & Pr($>$$|$t$|$) \\

##    \hline

## Intercept & 0.73 & 0.02 & 31.62 & 34.74 & 0.000 \\

##    Task & -0.02 & 0.02 & 17.29 & -0.87 & 0.397 \\

##    Age & 0.17 & 0.02 & 373.04 & 11.21 & 0.000 \\

##    Site & 0.05 & 0.02 & 385.12 & 2.39 & 0.017 \\

##    Task*Age & -0.02 & 0.01 & 2622.52 & -1.47 & 0.141 \\

##    Task*Site & -0.00 & 0.02 & 2631.48 & -0.20 & 0.842 \\

##    Age*Site & -0.00 & 0.02 & 388.66 & -0.06 & 0.951 \\

##    Task*Age*Site & 0.00 & 0.02 & 2639.46 & 0.21 & 0.834 \\

##     \hline

## \end{tabular}

## \caption{Model coefficients from a  linear mixed mode predicting the recognizability

## \end{table}


## % latex table generated in R 3.6.1 by xtable 1.8-4 package

## % Wed Jun 23 21:21:29 2021

## \begin{table}[ht]

## \centering

## \begin{tabular}{rrrrrr}

##    \hline

##  & Estimate & Std. Error & df & t value & Pr($>$$|$t$|$) \\

##    \hline
```

```
## Intercept & 0.72 & 0.02 & 34.02 & 37.49 & 0.000 \\

##    Task & -0.02 & 0.02 & 17.55 & -1.19 & 0.249 \\

##    Age & 0.14 & 0.02 & 376.68 & 9.30 & 0.000 \\

##    Site & 0.07 & 0.02 & 459.46 & 3.58 & 0.000 \\

##    Est. tracing score & 0.06 & 0.01 & 249.21 & 5.71 & 0.000 \\

##    Avg. intensity & 0.01 & 0.01 & 2839.40 & 1.90 & 0.058 \\

##    Draw duration & -0.03 & 0.01 & 1692.69 & -3.42 & 0.001 \\

##    Number of strokes & 0.03 & 0.01 & 2655.21 & 4.67 & 0.000 \\

##    Task*Age & -0.02 & 0.01 & 2623.63 & -1.72 & 0.086 \\

##    Task\&Site & -0.01 & 0.02 & 2630.59 & -0.45 & 0.655 \\

##    Age*Site & -0.01 & 0.02 & 415.59 & -0.74 & 0.462 \\

##    Task*Age*Site & 0.00 & 0.02 & 2638.72 & 0.18 & 0.858 \\

##     \hline

## \end{tabular}

## \caption{Model coefficients from a linear mixed mode predicting the recognizability o

## \end{table}


## Warning: Removed 18 rows containing missing values (geom_point).
```

Discussion

Acknowledgements

References

Barrett, M., & Light, P. (1976). Symbolism and intellectual realism in children's drawings. *British Journal of Educational Psychology*, *46*(2), 198–202.

Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychological Science*, *9*(3), 200–204.

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.

Cohn, N. (2012). Explaining "i can't draw": Parallels between the structure and development of language and drawing. *Human Development*, *55*(4), 167–192.

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.

Fan, J. E., Wammes, J. D., Gunn, J. B., Yamins, D. L., Norman, K. A., & Turk-Browne, N. B. (2020). Relating visual production and recognition of objects in human visual cortex. *Journal of Neuroscience*, *40*(8), 1710–1721.

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

Fury, G., Carlson, E. A., & Sroufe, A. (1997). Children's representations of attachment relationships in family drawings. *Child Development*, *68*(6), 1154–1164.

Hoffmann, D. L., Standish, C. D., Garcia-Diez, M., Pettitt, P. B., Milton, J., Zilhão, J., ... others. (2018). U-th dating of carbonate crusts reveals neandertal origin of iberian

cave art. *Science, 359*(6378), 912–915.

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.

Karmiloff-Smith, A. (1990). Constraints on representational change: Evidence from children's drawing. *Cognition, 34*(1), 57–83.

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

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.

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.

Long, B., Fan, J., & Frank, M. C. (2018). Drawings as a window into developmental changes in object representations. In *Proceedings of the 40th annual meeting of the cognitive science society.*

Minsky, M., & Papert, S. (1972). *Artificial intelligence progress report.* Cambridge, MA, USA: Massachusetts Institute of Technology.

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. *Journal of Neuroscience, 36*(42), 10893–10907.

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.

Piaget, J. (1929). The child's concept of the world. *Londres, Routldge & Kegan Paul.*

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.

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

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.

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.
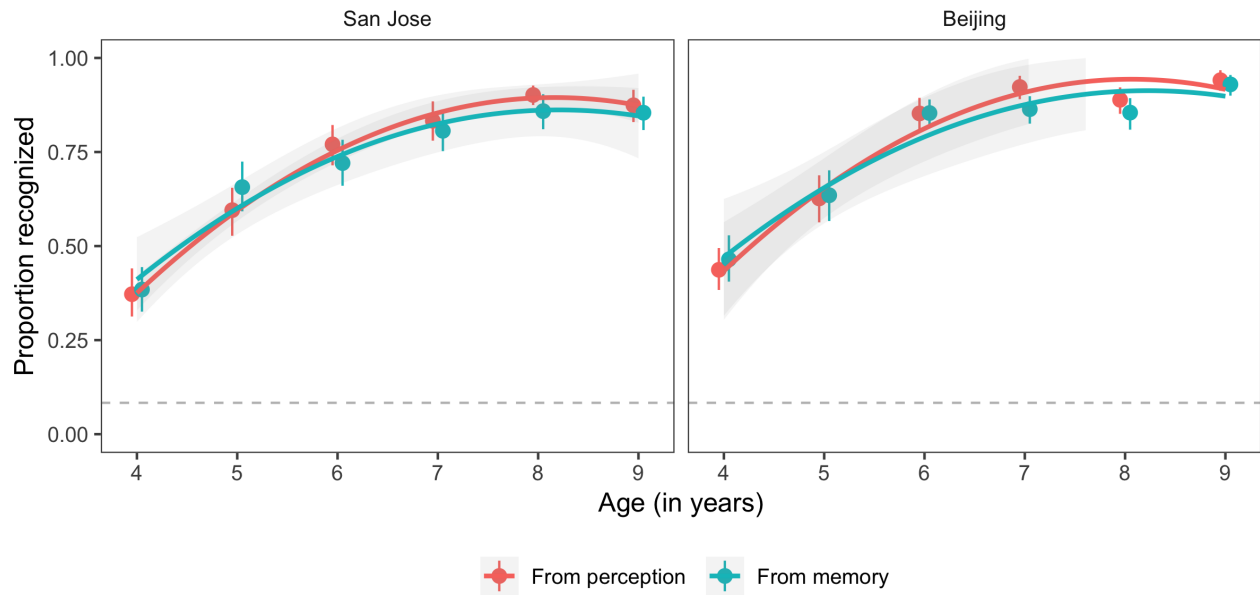
Figure 1. Proportion of drawings recognized as a function of the age of the child who completed each drawing, the geographical site they were tested at (San Jose vs. Beijing), and the type of drawing task they completed. Error bars show bootstrapped 95 percent cofnidence intervals.
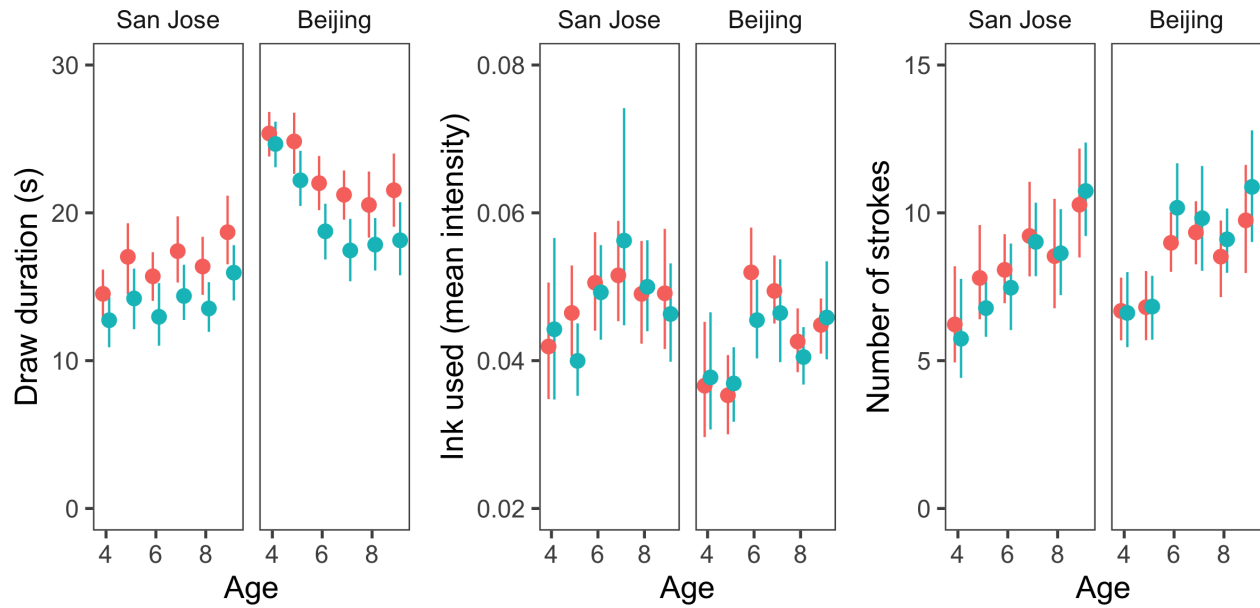
Figure 2. Effort covariates measured during the drawing task – amount of time spent drawing, amount of 'ink' used, and number of strokes used – as function of the age of the child who completed each drawing, the geographical site they were tested at (San Jose vs. Beijing), and the type of drawing task they completed. Error bars show bootstrapped 95 percent cofnidence intervals.
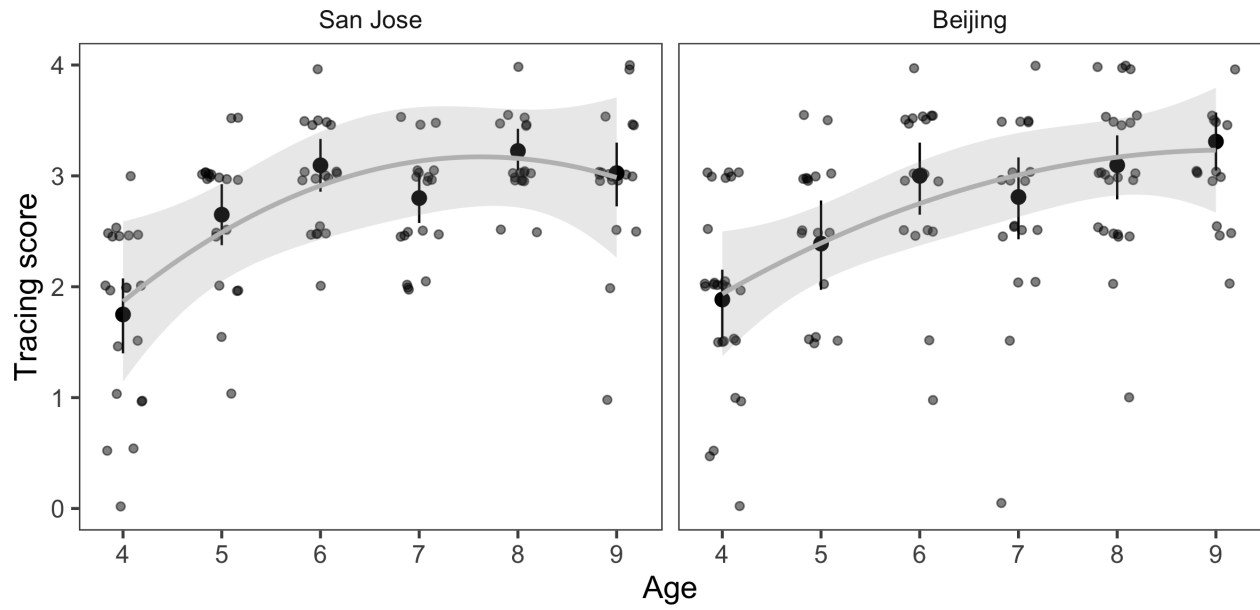
Figure 3. Average tracing scores across age and site; each dot represents an average tracing score obtained for each participant and are slightly jittered. Error bars represent bootstrapped 95 percent confidence intervals.