Drawings as a window into object representations in childhood

Bria Long
bria@stanford.edu
Department of Psychology
Stanford University

Judith E. Fan
jefan@stanford.edu
Department of Psychology
Stanford University

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

Abstract

A few well placed strokes can convey the identity of a person, object, or place. How do we become so skilled at creating graphical abstractions? Here, we take a developmental approach to question by asking children (ages 3-10 years) to draw a wide range of objects (e.g. 'can you draw a chair?') in under 30 seconds. Then, we analyze how the content of these drawings changes across development. We analyzed the degree to which children were able to depict recognizable exemplars of these categories, and found that children become increasingly skilled at producing recognizable drawings, even when controlling for low-level covariates (the amount of ink used, the time spent drawing, and the number of strokes). Second, we analyzed the degree to which children's drawings share perceptual features with those produced by adults, analyzing their featural similarity at each layer of deep convolutional neural network (VGG-19). Overall, we find that childrens' drawings were most similar to adults' in higher-level layers of the network. These results suggest that the ability to produce abstract graphical represnetations of objects is highly developed by middle childhood, and that even the most primitive drawings made by children contain featural similarities to the objects they are trying to depict. Broadly, these results point towards drawing as a promising avenue for investigating object representations in childhood.

Keywords: object representations; drawings; child development

Introduction

Consider what one has to do in order to draw "a phone" – one needs to access the mental representation of "a phone", distill this into a pictoral format, and plan a sequence of hand movements to effectively convey this visual concept. Yet this is a trivial task for ordinary adults. How do we learn to so effectively produce recognizable drawings? And might drawings offer a window into how young children represent common object categories?

While drawing has been extensively studied in early childhood, a primary focus has been on when children come to treat drawings as symbols for object categories (Gardner, 1994). And a wealth of evidence now suggests that in fact young children attribute rich meanings to their drawings. For example, children will attribute different symbolic content (e.g., "a balloon", "a lollipop") to very similar drawings based on what they intended to draw (Bloom & Markson, 1998). Further, children will monitor whether their drawings are adequate symbols for the things they are trying to draw, and will improve their drawings when given feedback that their drawings are not effective at communicating the identity of object (Callaghan, 1999).

Far less research, however, has examined if children's drawings reflect how children represent objects in the world around them. Indeed, drawings are a powerful way to tap internal representations of object categories, even in non-expert adults. For example, adults tend to draw objects that are small in the real-world at small visual sizes, and objects that are big in the real-world at big visual sizes (Konkle & Oliva (2011)). Further, to be recognizable, drawings have to depict the necessary features to express a given visual concept. This intuition is supported by recent computational work: deep neural network models of the ventral stream trained purely on photographs can also recognize drawings by non-expert adults, as drawings and photographs generate similar representations in higher-level layers of these models. In other words, drawings capture high-level similarity relationships between object categories (Fan, Yamins, & Turk-Browne, 2015).

Here, we explore how children draw common object categories across early childhood. First, we ask if children produce more recognizable drawings as they get older after factoring out low-level covariates. Second, we examine how similar children and adult's drawings are in terms of their perceptual features by comparing their representations in a deep convolutional neural network (VGG-19).

Part 1: Why do children get better at drawing?

First, children (ages 3-10 years) produced drawings of 16 common object categories in a simple drawing game. Then, nave adults attempted to recognize these drawings in a forced-choice recognition task.

Methods

Participants For the drawing task, children (N = 41, M = 6.9 years, range 4-10 years) were recruited at the San Jose Children's Discovery Museum and participated in this experiment. For the recognizability experiment, 14 adults with US IP addresses were recruited and rated all of the 268 drawings.

Materials We implemented a simple drawing game in HTML/Javascript using the paper.js library; this web-based experiment was run on an iPad on the floor of the museum. All code is available at www.github.com/brialorelle/kiddraw/museumdraw.

Drawing Game Procedure On each trial, a text cue would appear (i.e., "Can you draw a [flower]?") that the experimenter would read out, ("What about a [flower]? Can you draw a [flower]?). Then, a drawing canvas appeared (600 x 600 pixels) and children were had 30 seconds to make a drawing before the game moved on to the next trial. After each trial, the experimenter asked the child whether they wanted to keep drawing or whether they were all done. On the first two trials of the experiment, every child was prompted to draw the same two common shapes—-a circle and a triangle. These tri-

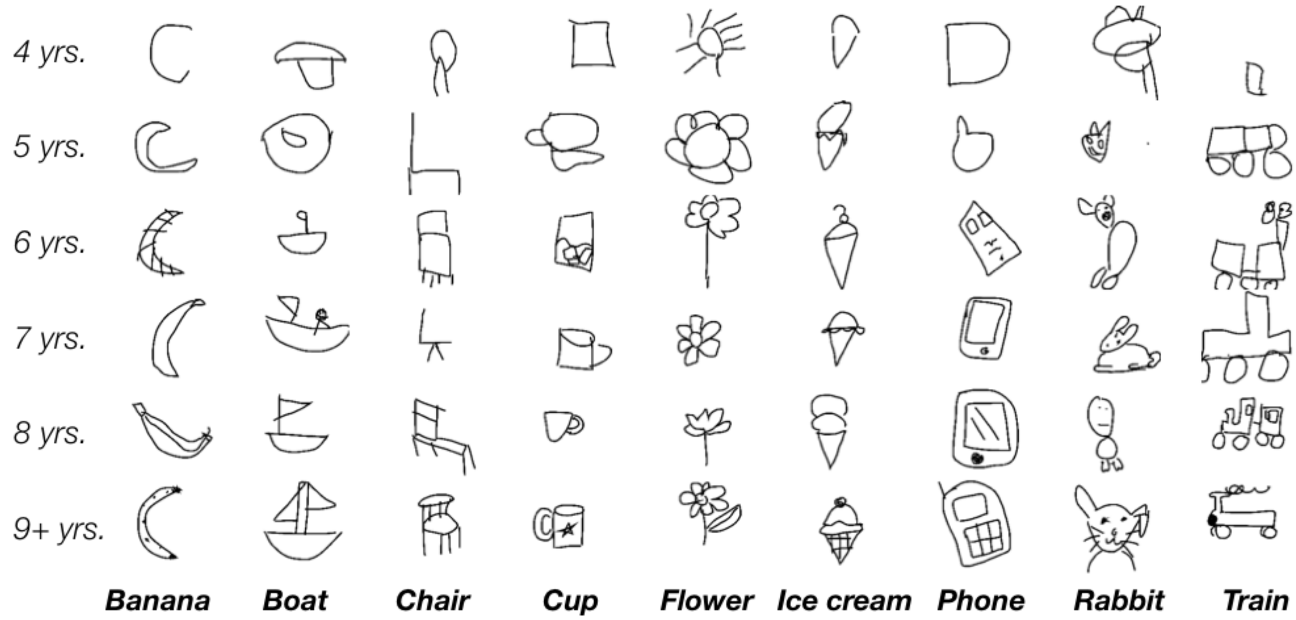

Figure 1: Example drawings made by children ages 4-10 of several object categories.

als served to familiarize children with the drawing task and to practice using their fingers to draw.

Stimuli Stimuli were words referring to 16 common object categories (banana, boat, car, carrot, cat, chair, couch, cup, flower, foot, frog, ice cream, phone, rabbit, shoe, train). These categories were chosen such that they were (1) likely to be familiar to children, (2) present in the Google QuickDraw database, (3) spanned the animate/inanimate distinction and (4) intuitively spanned a wide range of difficulty (for example, flowers seem easier to draw than couches).

Recognizability Task 14 nave adults assessed the recognizability of all of the 286 drawings produced by these children. On each trial, participants saw a drawing, and were asked "What does this look like?", and responded by typing into a text box; participants could then choose between 21 possible answers. 16 of these possible answers were the original object categories; however, we also included five additional foil items (bean, arm, person, rock, and "cannot tell at all"). All drawings were presented in a random order, and participants were not informed that these drawings were produced by children or the context in which they were produced. An answer was scored as "correct" if adults were able to correctly guess the object category that children were cued with.

Low-level covariates. The use of a digital interface for drawing allowed us to quickly and easily assess the contribution of several low-level factors that may co-vary with drawing ability. For each drawing, we thus quantified the amount of time spend drawing, the number of strokes used, and the overall intensity of the drawing(e.g., amount of ink). Descriptives plots describing the output of these variables can be seen in (see Figure 2, left).

GLMM procedure. We aimed to assess whether children's ability to produce recognizable drawings increased with age, independent of low-level covariates. To do so, we used a generalized logistic mixed effect model, with age, drawing duration, amount of ink used, and number of strokes as fixed effects, and with random effects for each individual child drawer and object category.

Results

First, we observed that some items were much easier to draw than others. For example, children of all ages produced drawings of cats that were readily recognizable as "cats", but few children of any age produced drawings that were recognizable as "shoes" (see Figure 3). Howver, almost all items also saw an increase in recognizability with the age of the drawer. Across all items, the proportion of drawings recognized increased steadily with age (% drawings recognized; chance = 4.8%; M_{4yrs} = 14%, M_{5yrs} = 45%, M_{6yrs} = 70%, M_{7yrs} = 72%, M_{8yrs} = 66%, M_{9yrs} = 76%, M_{10yrs} = 85%).

Next, we asked whether this relationsip persists when we control for low-level covariates: the number of strokes, amount of ink used, and the time spent drawing. In other words, is this increase in recognizability due to an increase in expressive power, or simply due to the fact that older children may have put more effort into their drawings? Our generalized logistic mixed-effect model revealed that the recognizability of drawings increased reliably with when controlling for these low-level covariates — the amount of time spent drawing, the number of strokes, and total ink used (b = 0.96,

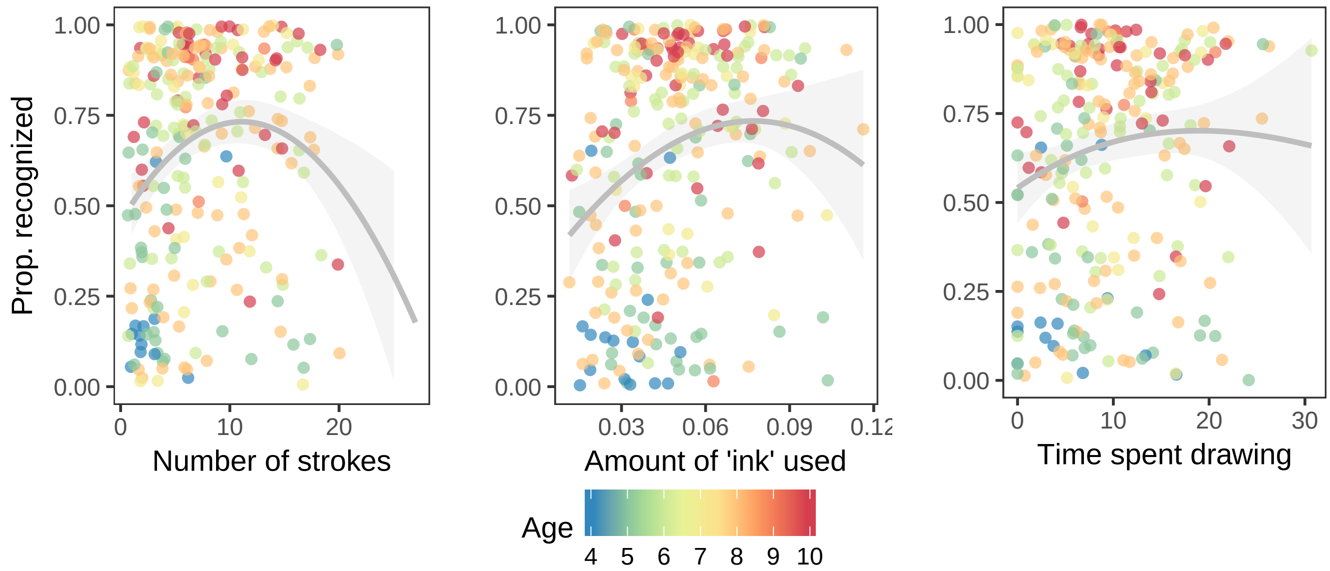

Figure 2: The proportion of adults who recognized each drawing is plotted as a function of the number of strokes, amount of ink used, and the time spent creating each drawing. Each dot represents an individual drawing; dots are colored by the age of the drawer.

SE = 0.17, Z = 5.5), and accounting for variation across object categories and individual children. All model coefficients can be seen in Table 1.. Adding interaction terms between age and these low-level covariates did little to decrease the effect of age on recognizability (b = 0.94, SE = 0.18, Z = 5.4).

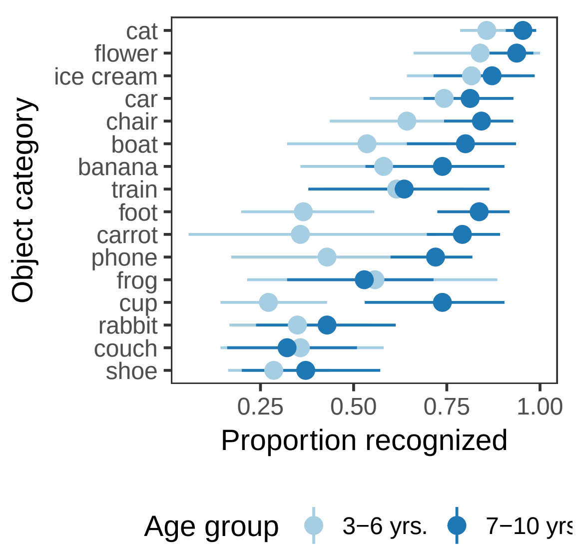

Figure 3: Proportion of drawings recognized for object category, sorted from hardest to easiest items. Error bars represent non-parametric 95 percent confidence intervals, estimated using the langcog r package.

Thus, these results suggest that the ability to quickly produce graphical representations of object categories increases with age, independently of low-level covariates. Further, these results sugest that this ability is highly developed by middle childhood, plateauing for these object cateogries around age 6-7.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.861	0.321	2.680	0.007
Age	0.956	0.174	5.497	0.000
Drawing time	0.338	0.109	3.105	0.002
Amount of ink	0.014	0.080	0.179	0.858
Num. strokes	-0.289	0.098	-2.959	0.003

Table 1: Model coefficients of a GLMM predicting the recognziability of each drawing.

Part 2: How similar are children's and adult's drawings?

To what degree are children's drawings similar to those of adults? While younger children often produced drawings that were unrecognizable at the basic-level, these recognizability ratings may underestimate the perceptual content depicted in children's drawings. For example, children may not be able to depict the visual differences between a bunny and a frog, but they still may capture many of the essential perceptual features needed to depict an animal. Here, we turn to deep neural network models of object recognition to quantify the similarities between children's and adults drawings, asking how similar they are in terms of feature similarity at each progressively complex layer of a deep convolutional neural network (Simonyan & Zisserman, 2014). In other words, how similar are children and adult's drawings (and why)?

For this analyses, we collected a larger sample of drawings using the same methodology, this time sampling from both the previously used categories as well as a new selection of 22 categories (see Stimuli) allowing us to span superordinate category distinctions. With this larger sample of drawings, we thus examine the degree to which children and adults drawings resemble each other in a deep convolutional neural network, specifically VGG-19.

Methods

Stimuli For this second round of data collection, we expanded our set of categories to include equal numbers of vehicles, furniture, small objects, food items, mammals, and non-mammals (new items: airplane, bus, bike, piano, table, door, bed, fork, keys, hat, apple, cookie, mushroom, horse, dog, sheep, bear, fish, bird, spider, shark, duck).

Participants Participants included those who participated in the first round of data collection, used in Experiment 1, as well as an additional 37 children, again recruited from the floor of the San Jose Children's Discovery Museum. Overall, this yielded an additional 98 drawings (excluding practice trials) for a total of 387 drawings. For the subsequent analyses, we binned children's age coarsely into "younger children" (aged 3-6 years) and "older" children (aged 7-10 years) and restricted analyses to categories where we had at least 3 drawings per age group (16 categories); this allowed us to analyze approximately the same number of drawings in each age group (younger children, N=118 drawings; older children, N=161 drawings). This ensured the robust estimates of feature distance in the folowing analyses.

Adult drawings We obtained a sample of adult drawings from the Google QuickDraw database. Specifically, we randomly sampled 1000 images for each object category, irrespective of any information about the adult drawer or the quality of the drawing. See `https://quickdraw.withgoogle.com/data` for visualizations of this dataset.

Convolutional Neural Network (CNN) Features We used a standard, pre-trained implementation of VGG-19 (Simonyan & Zisserman, 2014) to extract features in response to all sketches at each layer of the network, including the first five convolutional layers (C1-C5) as well as the two fully-connected layers (FC6 and FC7). Features were normalized within each layer across all sketches and then averaged within each category (e.g., "cat", "rabbit"). This yielded a vector corresponding to the number of features in each layer for all 38 of the drawn categories in younger children, older children, and adults.

Representational Similarity Analyses Separately for drawings from younger children, older children, and adults, we averaged the feature vectors within each object class for a given layer of VGG and then computed a layer-specific matrix of the Pearson correlation distances between these average vectors across classes (Kriegeskorte et al., 2008). Formally, this entailed computing:

$$RDM(R)_{ij} = 1 - \frac{cov(\vec{r}_i, \vec{r}_j)}{\sqrt{var(\vec{r}_i) \cdot var(\vec{r}_j)}}$$

, where $\vec{r}_i$ and $\vec{r}_j$ are the mean feature vectors for the ith and jth object classes, respectively. Each of these 16x16 representational dissimilarity matrices (RDMs, shown in Figure 4) provides a compact description of the layout of objects in the high-dimensional feature space inherent to each layer of the model. Following Kriegeskorte et al. (2008), we measured the similarity between object representations in different layers by computing the Spearman rank correlations between the RDMs for those corresponding layers.

Estimates of standard error for the Spearman correlation between RDMs (i.e., between domains or between layers) were generated by jackknife resampling of the 16 object classes. This entails iterating through each of the 16 subsamples that exclude a single class, computing the correlation on each iteration, then aggregating these values. Specifically, the jackknife estimate of the standard error can be computed as: $s.e._{(jackknife)} = \sqrt{\frac{n-1}{n}\sum_{i=1}^{n}(\bar{x}_i - \bar{x}_{(.)})^2}$, where $\bar{x}_i$ is the correlation based on leaving out the ith object class and $\bar{x}_{(.)} = \frac{1}{n}\sum_i^n \bar{x}_i$, the mean correlation across all subsamples (of size 15). This estimate of standard error allows us to construct 95% confidence intervals and compute two-sided p-values for specific comparisons (Efron, 1979; Tukey, 1958).

Category similarity analyses We also directly analyzed the similarity of the feature representations generated by sketches from younger children vs. adults and between older children vs. adults. To do so, we computed the Pearson correlation between the average category vectors for adults and the average category vectors for younger/older children (separetly). This yielded a correlation score for each object category for each age comparison.

Category classification analyses Model features were also used to train softmax classifiers (`http://scikit-learn.org/`) with L2 regularization to evaluate the degree to which category information was linearly accessible from sketches made by each group of participants. Predictions are then made for images held out from the training set, and accuracy is assessed on these held-out images. The robustness of classifier accuracy scores was determined using stratified 5-fold cross validation on 80% train/20% test class-balanced splits.

Results

Layer-wise feature similarity We first examined the featural similarities between sketches produced by adults and children at each layer of VGG-19. Overall, we found that the similarity between older children and adults' drawings increased in each subsequent layer of the network, reaching a peak in the final layers of the network (see Figure 5; Spearman's r values, Layer 1=0.215, Layer 2=0.206, Layer 3=0.249, Layer 4=0.435, Layer 5=0.637, Layer 6=0.696, Layer 7=0.72). For younger children, we found a similar pattern of results, though similarity to adult drawings was overall lower (Spearman's r values, Layer 1=0.074, Layer 2=0.001, Layer 3=0.04 , Layer 4=0.119, Layer 5=0.368, Layer 6=0.464, Layer 7=0.49). The RDMs for the final layer

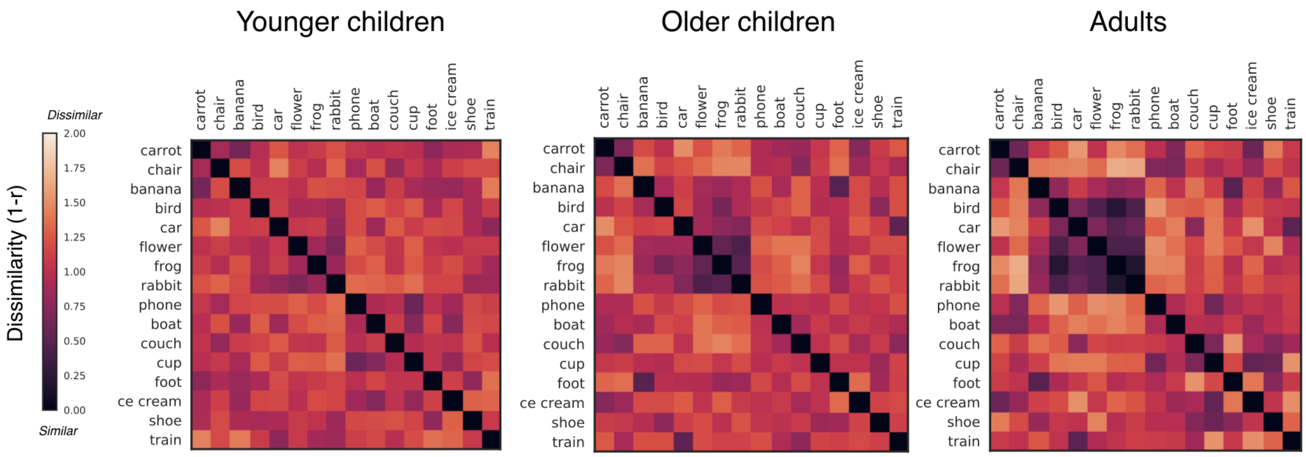

Figure 4: Representational dissimilarity matrixes (RDMs) in the highest layer of VGG-19 (FC7) for drawings made by younger children (3-6 years of age), older children (7-10 years of age), and adults (Google QuickDraw database). Each square in one of these matrixes represnets the correlation distance between two categories (e.g., chair and couch) in this layer of the network; lighter colors indicate pairs of categories that generated dissimilar feature representations; darker colors indicate pairs of cateogries that generated more similar feature representations. Categories are grouped to reveal the inherent simiarity structure.

of the network (where similarity was the highest; FC7) are shown in Figure 4 for younger children, older children, and adults.

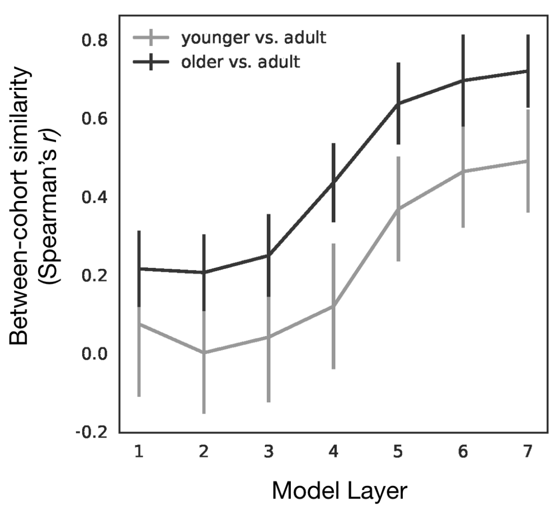

Figure 5: Spearman's correlation between representational dissimilarity matrixes (RDMs) between drawings produced by adults and younger children (grey line) and between adults and older children (black line) at each layer of VGG-19–the first five convolutional layers and the last two fully connected layers. Error bars represent standard error of the mean obtained by a jacknife re-sampling procedure (see Methods).

Category similarity analyses Next, we explored which categories generated representations in FC7 that were more or less similar between younger children vs. adults and between older children vs. adults (see Figure 6). Overall, we found a good deal of variability; for some categories, children's and adults drawings had feature representatinos that were relatively dissimilar (e.g., couches, shoes) while others generated very similar feature representations (e.g., flowers, chairs) in this final layer of the network.

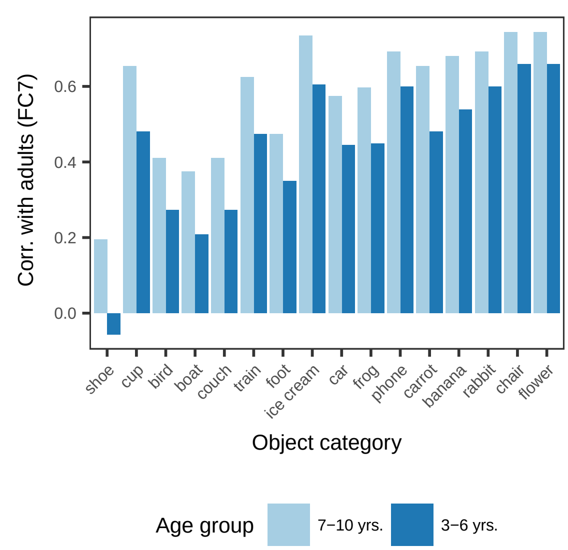

Figure 6: Spearman's correlation between children's and adults sketches in layer FC7 for each object category.

Classification results Finaly, we examined the degree to which these featural representations could be used to classify these skeches at the basic-level. We found that sketches made by younger children were classifiable 35% of the time

(SD=5%), while those made by older children (7-10 years) were classifiable 51% of the time (SD=6%). While the overall performance of the classified is relatively low compared to the human performance seen in Part 1, we still observed a relative increase in recognizability between younger and older children. Thus, these results suggest that the difference in recognizability of the sketches stems directly from a differences in perceptual features that can be detected by a deep convolutional neural network trained to recognize objects.

Taken together, these results suggest that children and adults are accessing similar category representations to perform these drawing task that manifest in perceptual similarities between adults and children, even in the most primitive drawings.

General Discussion

We explored how children get better at drawing and how similar their drawings are to adults'. Overall, we found that the capacity to quickly produce graphical representations that communicate object category information is highly developed by middle childhood. Children produced drawings in under 30 seconds that were recognizable by adults with only the few "strokes" of a pen; older children were better able to produce these graphical abstractions than younger children, irrespective of the amount of time they spent drawing or the amount of "ink" they used. Further, children's drawings were most similar to adult drawings in the higher-level layers of a deep convolutional neural network trained to recognize objects, suggesting that children and adults drawings share high-level perceptual features useful for basic-level object recognition.

An obvious future direction concerns the contribution of children's motor control to their drawing abilities. In other words, to what degree are drawings made by older children more recognizable simply because older children have better fine motor control? Children certainly practice drawing—-both on their own and in structured settings (e.g., art classes)–and this practice invariably plays a role in the fidelity of drawings that children can produce. We plan to measure children's fine motor control on an orthogonal task (e.g., tracing a complex shape) to begin to understand how this factor influences the recognizability of children's drawings.

Ultimately, we hope to understand the degree to which changes in children's drawings of object categories reflect changes in children's representations of object categories. Throughout childhood, children certainly acquire a wealth of experience with the objects in the world around them, and this experience likely helps build more detailed internal representations of the categories. Thus, one possibility is that children's internal representations that they map to the words "rabbit", "chair", and "couch" are becoming more detailed as they grow older, and that it is these more detailed representations that, in turn, feed into their rapid drawings of these object categories. If this is the case, then children's abilities to depict certain object categories may patterns with their object categorization errors. For example, older (vs. younger) children may be better able to draw cats vs. rabbits and better able to distinguish between cats vs. rabbits. Future work that links children's categorization abilities with their visual production behaviors may begin to answer this question.

In sum, this work begins a developmental project examining children's object representations using a common production task–drawing. An understanding of how we produce efficient graphical abstractions of the objects in our everyday world may help uncover the building blocks of our object category representations.

Acknowledgements

We gratefully acknowledge those who made the Google QuickDraw database avaliable. This work was funded by a ## to Judy Fan, and a ### to Michael C. Frank, and a ### to Bria Long.

References

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

Callaghan, T. C. (1999). Early understanding and production of graphic symbols. *Child Development*, *70*(6), 1314–1324.

Efron, B. (1979). 1977 Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife. *Annals of Statistics*, *7*(1), 1–26.

Fan, J. E., Yamins, D., & Turk-Browne, N. B. (2015). Common object representations for visual recognition and production. In *CogSci*.

Gardner, H. (1994). *The arts and human development: With a new introduction by the author*. Basic Books.

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 23.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv Preprint ArXiv:1409.1556*.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics*.