Drawings as a window into developmental changes in object representations

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

How do children's representations of object categories change as they grow older? As they learn about the world around them, they also express what they know in the drawings they make. Here, we examine drawings as a window into how children represent familiar object categories, and how this changes across childhood. We asked children (age 3-10 years) to draw familiar objects categories on an iPad. First, we analyzed their semantic content, finding large and consistent gains in how well children could produce drawings that are recognizable to adults. Second, we quantified their perceptual similarity to adult drawings using a pre-trained deep convolutional neural network, allowing us to visualize the representational layout of object categories across age groups using a common feature basis. We found that the organization of object categories in older children's drawings were more similar to that of adults than younger children's drawings. This correspondence was especially strong for higher layers of the neural network, showing that older children's drawings tend to capture high-level perceptual features critical for adult recognition. We hypothesize that this improvement reflects increasing convergence between children's representation of object categories and that of adults; future work will examines how these age-related changes relate to children's emerging perceptual and motoric capacities. Broadly, these findings point to drawing as a rich source of insight into how children represent object concepts.

Keywords: object representations; drawings; child development

Introduction

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 at least 40,000 years (Pike et al., 2012), 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, providing a rich source of potential insight into their emerging understanding of the visual world. For example, as children learn the diagnostic properties of objects they encounter, they might express this knowledge in the drawings they make. How can we leverage this natural behavior to understand how they learn abstractions over their perceptual experience, such as object categories?

On the one hand, children quickly form sophisticated perceptual representations of familiar objects, leveraging shape information in conjunction with linguistic cues (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). Typically, such learning is measured using discrete choices between stimuli that vary along dimensions chosen by an experimenter, and large numbers of discrimination trials are required to yield reliable estimates of perceptual performance. By contrast, drawing tasks both permit children to include any information they consider relevant and can provide rich, high-dimensional information about the content and structure of children's perceptual representations. For example, when presented with a target object to draw in which a prominent feature is occluded (e.g., the handle of a mug is turned away), children as young as 5 years of age frequently include the occluded object part in their drawing anyway, displaying the robustness of their internal representation to variation in viewpoint (Davis, 1983).

On the other hand, important developmental changes in perceptual processing continue throughout childhood (Juttner, Wakui, Petters, & Davidoff, 2016; for reviews, see Nishimura, Scherf, & Behrmann, 2009). For example, young children tend to categorize novel objects on the basis of part-specific information, whereas older children additionally recruit information about relationships between object parts (Mash, 2006). This set of changes is also resonant in children's drawings: there appear to be dramatic changes in how children encode semantically relevant 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) or relational (Light & Simmons, 1983) information.

But while figurative drawings have long provided inspiration for scientists investigating the representation of object concepts in early life (Minsky & Papert, 1972), a major barrier has been the lack of principled quantitative measures of high-level perceptual information in drawings. As such, previous studies employing drawing tasks have typically relied on qualitative assessments (Kosslyn, Heldmeyer, & Locklear, 1977) or ad hoc quantitative criteria (Goodenough, 1963). Recent work in computational vision has validated the use of pre-trained deep convolutional neural network (DCNN) models to quantitatively measure high-level perceptual information in adult drawings (Fan, Yamins, & Turk-Browne, 2015). Higher layers of these models both capture adult perceptual judgments of object shape similarity (Kubilius, Bracci, & Op de Beeck, 2016) and predict neural population responses in categories throughout object-selective cortex (Yamins et al., 2014). Thus, features learned by these models provide a principled choice of a basis for extracting semantic information regarding object identity from children's drawings.

Here we examine children's drawings as a window into how they represent familiar visual object categories, and how

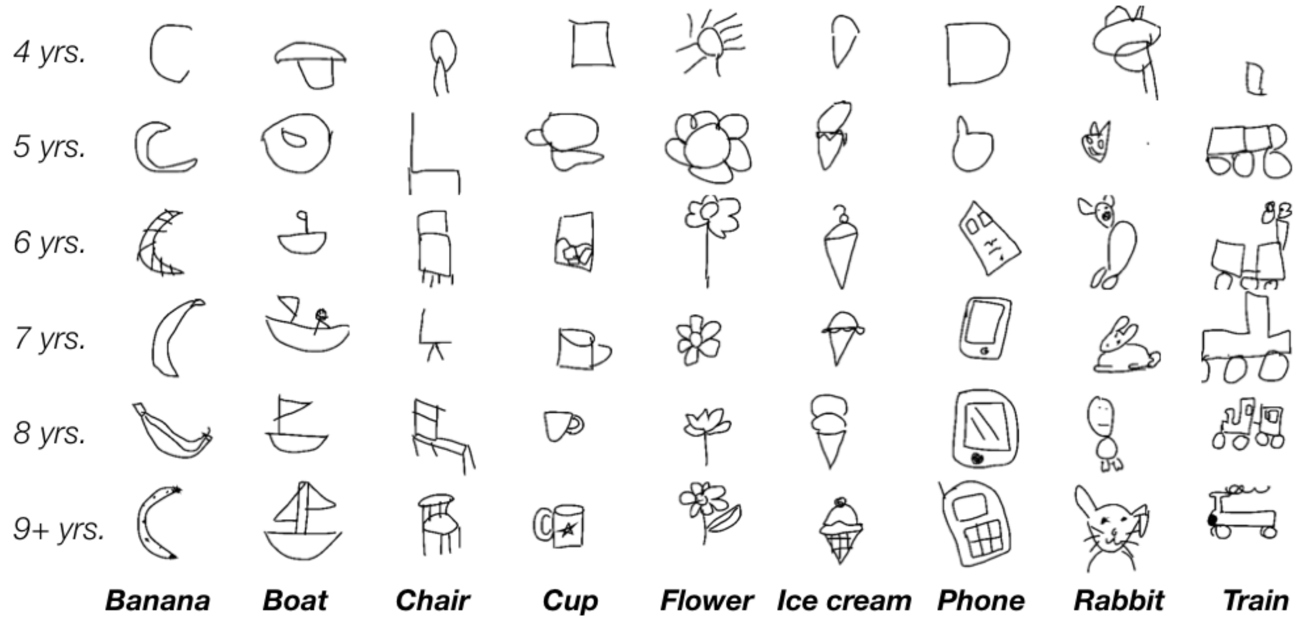

Figure 1: Example drawings made by children ages 4-10 of several object categories.

this representation and its translation into graphical format changes across childhood. To do so, we asked children (ages 3-10 years) to draw a variety of object categories on a digital tablet. Afterwards, adults attempted to recognize these drawings in a forced-choice recognition task. In Part 1, we examine how this semantic information in children's drawings changes with age by after factoring out low-level covariates related to motor production, such as how long they spend drawing and how many strokes they use. In Part 2, we compare the high-level perceptual features of child and adult drawings by relating their representations in a pre-trained DCNN model, allowing us to visualize the representational layout of object categories across age groups using a common feature basis.

Part 1: Semantic information in children's drawings

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. Either the child or their parents verbally reported the child's age. For the recognizability experiment, 14 nave adults with US IP addresses were recruited from Amazon Mechanical Turk and provided labels for all drawings.

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) spanned the animate/inanimate distinction, and (3) intuitively spanned a wide range of difficulty (for example, flowers seem easier to draw than couches). We also choose items that were present in the Google QuickDraw database, a large dataset containing drawings made by adults in under 20 seconds, so that we could eventually compare children's drawings with ones made by adults.

Drawing Task Procedure We implemented a web-based drawing game in HTML/Javascript using the paper.js library and collected drawings using a touchscreen tablet on the floor of the museum. At the beginning of each session, to familiarize children with the task and touch interface, they were prompted to draw a circle and a triangle. After completing these two practice trials, they were cued to draw a randomly selected object. 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 had 30 seconds to make a drawing before moving onto the next trial; pilot testing suggested that 30 seconds was enough for many children to complete their drawings. After each trial, the experimenter asked the child whether they wanted to keep drawing or whether they were all done. In all, we collected 268 drawings across the 16 categories.

Recognizability Task Procedure After collecting children's drawings, we presented them to nave adults to measure their recognizability. On each trial, participants saw a drawing, and were asked "What does this look like?", and responded by typing their response into a text box. Only labels from a restricted set of 21 options were accepted, comprising the 16 drawn categories, 4 foil categories (bean, arm, person, rock), and "cannot tell at 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.

Model Fitting Our goal was to measure how children's ability to convey semantically relevant information in their drawings changes with age. We anticipated that their drawings may also vary along other dimensions more directly related to the motor production demands of the task, such as the amount of time spent drawing, the number of strokes used, and amount of ink (i.e., mean pixel intensity of sketch).

In order to assess whether children's ability to produce recognizable drawings increased with age, independent of these low-level covariates, we fit a generalized linear mixed-effects model, with scaled age (specified in years), drawing duration, amount of ink used, and number of strokes as fixed effects, and with random intercepts for each individual child and object category. The dependent variable was whether the proportion of adults that recognized a given drawing.

	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 recognizability of each drawing.

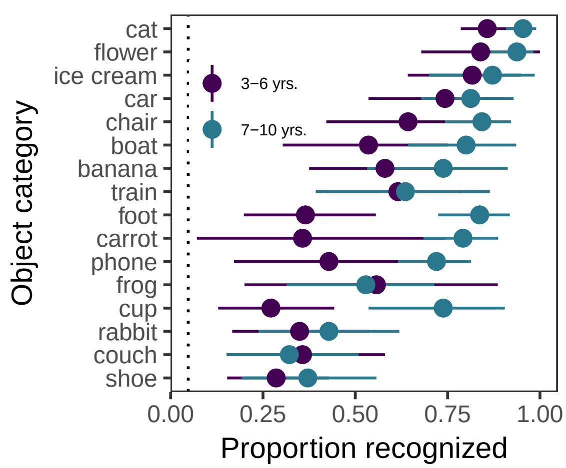

Figure 2: Proportion of drawings recognized in each object category. The dashed line represents chance performance. Error bars represent non-parametric 95 % confidence intervals.

Results

We found that recognizability of drawings generally increased with age (see Figure 3), although there was substantial variability across categories in how well children could produce recognizable drawings. 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 2).

Was this increase due to an increase in high-level semantic properties of older children's drawings, or to the possibility that older children may have put more time and effort into their drawings? Our mixed-effects model revealed that recognizability of drawings reliably increased when controlling for these low-level covariates — the amount of time spent drawing, the number of strokes, and total ink used ($\beta = 0.96$, $SE = 0.17$, $Z = 5.5$), 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 ($\beta = 0.94$, $SE = 0.18$, $Z = 5.4$).

Taken together, these results show large and consistent gains in how well children can produce recognizable drawings across this age range, although younger children still produced drawings that could be recognized well above chance by adult viewers.

Part 2: Perceptual information in children's drawings

In the previous section, we found that children's drawings generally contained sufficient semantic information to support recognition by adult viewers, although older children's drawings were consistently more recognizable. What is the nature of the developmental change that underlies older children's enhanced ability to produce recognizable drawings (at least to adult viewers)? And how might children's drawings provide a window into their perceptual representations of these objects?

We hypothesized that this improvement reflects an increasing convergence in the perceptual features depicted in children and adult's drawings of their object representations. To measure this, we extracted the high-level perceptual features of drawings made by children and adults using a pre-trained deep convolutional neural network (Simonyan & Zisserman, 2014). These features form a common basis for representing shape similarity – including the presence of diagnostic object parts (e.g., legs, handles) – and a basis from which object identity can be easily derived (Kubilius et al., 2016). We then use these high-level features to evaluate how similar the representational layout of object categories is between children and adult's drawings. Insofar as they are more similar for older children than younger children, this could explain why adults are more accurate in recognizing their drawings.

To evaluate whether these high-level features are necessary to capture semantically-relevant perceptual information, we also analyze these drawings using features from earlier layers of the model. In DCNN models, the earliest layers tend to capture local image statistics, including edges and textures, whereas intermediate layers capture mid-level features, including curvature (Yamins et al., 2014). While the earliest layers are sensitive to the spatial location of certain visual

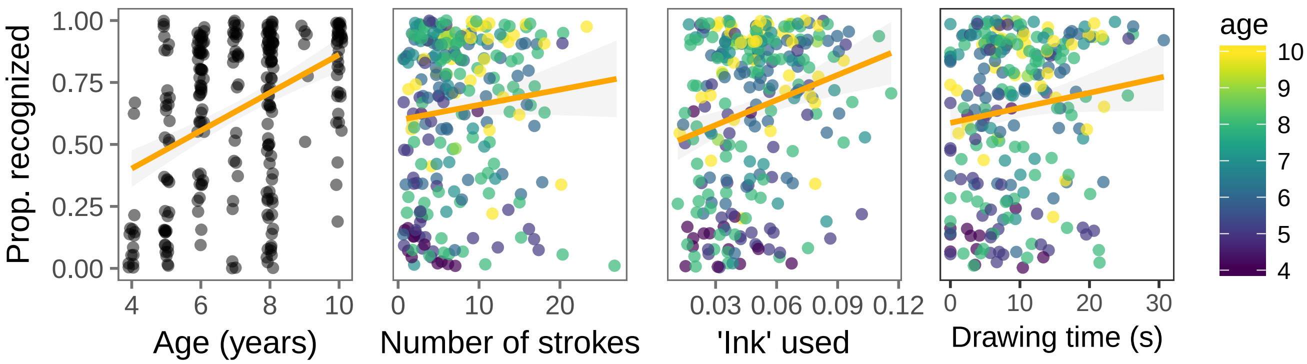

Figure 3: The proportion of adults who recognized each drawing is plotted as a function of child's age, the number of strokes, amount of ink used, and the time spent creating each drawing. Each dot represents an individual drawing; dots in the right three plots are colored by the age of the drawer.

features (e.g., whether there are horizontal lines on bottom vs. the top of the canvas), later layers compute representations that are increasing invariant to spatial position. Thus, if adults and children's drawings are most similar in terms of these more basic visual features, similarity should instead peak in these earlier layers.

Methods

Participants Participants included those who participated in the first round of data collection, as well as an additional 37 children recruited in the same way as in Part 1.

Drawing dataset In our second round of data collection, our goal was to expand the number of categories included in our model-based feature analyses, so we included an additional 22 categories. Across both rounds of data collection, we recorded 374 drawings from 78 children across a broad age range. However, due to the limited amount of data in each category for each age, in subsequent analyses we group drawing data into two coarse age categories: younger children (aged 3-6 years) and older children (aged 7-10 years). We thus restricted the following analyses to the 15 categories where we had at least 3 drawings in both younger and older age groups, yielding 113 drawings by younger children and 157 drawings by older children. Including a minimum number of drawings per class and age category ensured robust estimates of category-level feature information in drawings.

To complement the children's drawing dataset, we obtained a random sample of 100 adult drawings from each of the categories above from the Google Quickdraw dataset (`https://quickdraw.withgoogle.com/data`). Prior to analysis, we cropped all sketch images to contain only the sketch, applied uniform padding (10px), and rescaled them to the same size (3x224x224).

Deep convolutional neural network model We used a standard, pre-trained implementation of the VGG-19 architecture (Simonyan & Zisserman, 2014) to extract features from sketches at layers across several depths in the network. Specifically, we analyzed feature activations in the first five pooling layers, as well as the first two fully-connected layers. Each image elicits a pattern of feature activations at every layer in the model, where each pattern is equivalent to a vector in a feature space with the same number of dimensions as units in that layer.

Representational Similarity Analyses Separately for the younger-children, older-children, and adult drawing datasets, we averaged the feature vectors within each object category in both pixel space and for a given layer of VGG-19 and then computed a layer-specific matrix of the Pearson correlation distances between these average vectors across categories (Kriegeskorte, Mur, & Bandettini, 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 categories, respectively, where R represents the correlation betwen two categories (e.g., rabbits and shoes). Each of these 15x15 representational dissimilarity matrices 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 15 object categories. 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). This entails iterating through each of the 15 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 $\bar{x}_{(.)} = \frac{1}{n}\sum_i^n \bar{x}_i$,

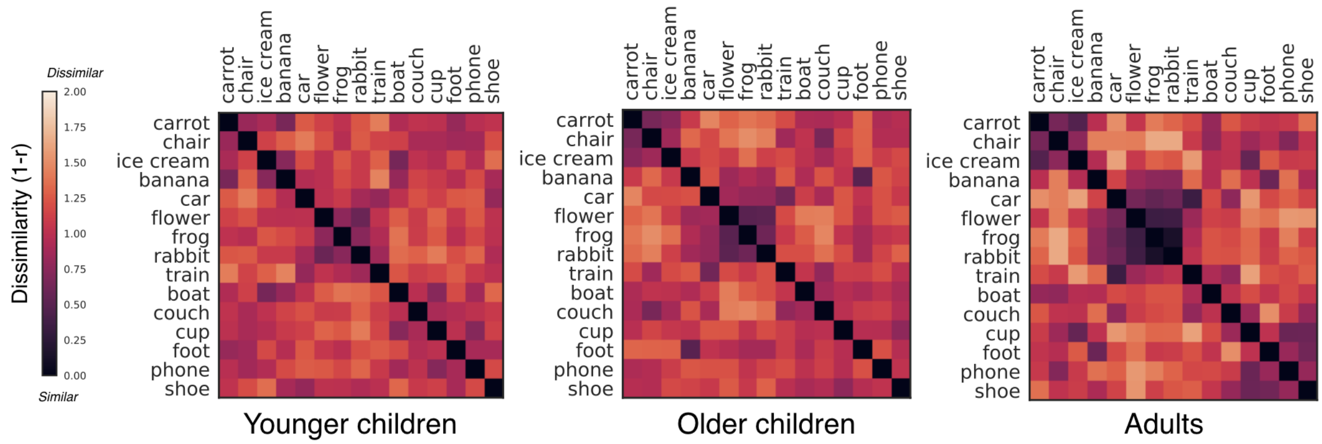

Figure 4: Representational dissimilarity matrices (RDMs) in the highest layer of VGG-19 (FC7) for drawings made by younger children (3-6 years), older children (7-10 years), and adults.

the mean correlation across all subsamples (of size 15).

Results

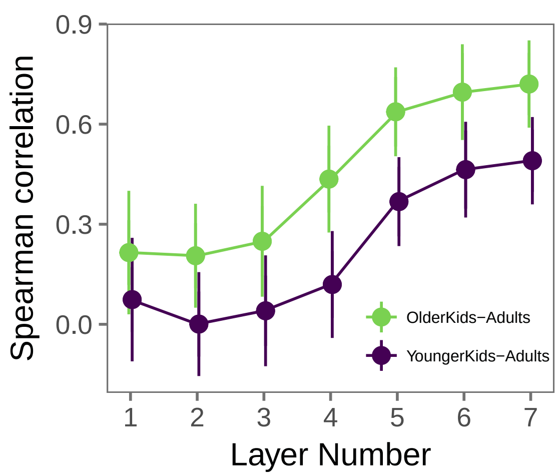

Figure 5: Spearman's correlation between representational dissimilarity matrices (RDMs) of drawings produced by adults vs. other adults, adults vs. older children, and between adults vs. younger children at each layer of VGG-19. Error bars represent standard error of the mean obtained by a jacknife re-sampling procedure (see Methods).

We examined the feature similarities between sketches produced by adults and children at each layer of VGG-19. For older children, 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). For younger children, we found a noisier pattern of results, as similarity to adult drawings was lower overall. Conversely, adult and children's drawings were dissimilar in pixel space for both age groups (adults vs. older children, r = .06; vs. younger children, r = .02).

The RDMs for the final layer of the network (where similarity was the highest; FC7) are shown in Figure 4. Here, each cell represents the correlation distance between two categories in this layer of the network; lighter colors indicate pairs of categories that are further apart in feature space; darker colors indicate pairs of categories that are closer. For visualization purposes, categories are ordered according to the clusters that appear in the data.

Inspecting the similarity structure for adults, drawings of animals (rabbits, frogs, and flowers) elicited the most similar representations, evident in a cluster in the middle of the RDM. While there are some intuitive category pairings (e.g., drawings of cars are somewhat similar to trains), other clusters seem less intuitive at first blush (carrot–chair–ice cream)—yet note that all three of these categories tend to have a similar aspect ratio. Importantly, this perceptual similarity structure was echoed in children's drawings of these same object categories. Children and adult's drawings appear to share many high-level perceptual features useful for object recognition.

General Discussion

What do children's drawings reveal about their object representations? We approached this question by analyzing the semantic and perceptual content of children's drawings across childhood. We found that the capacity to quickly produce drawings that communicate category information improves with age, even when factoring out low-level motor covariates. In addition, we found that drawings from older vs. younger children were more similar to adult drawings in high-level layers of a deep convolutional neural network trained to recognize objects, suggesting that older children's drawings also contain more of the perceptual features relevant for recognition. Children and adults may be accessing similar category representations when asked to "draw a chair" and communicating this representation through their drawings.

A natural question is how any age-related differences in drawings are related to children's ability to control and plan their hand movements. While drawing recognizability increased with age when accounting for the low-level motor covariates, these measures likely only partially estimate children's motor abilities. We plan to measure both children and adult's ability to perform orthogonal fine motor tasks (e.g., tracing a complex shape) to understand how motor developments influence the drawings that children produce.

At the same time, children are also continuously learning about new object categories and their properties. How might this learning affect children's internal representations (and drawings) of different object categories? One possibility is that the bulk of the development change revolves around building more detailed representations: children may be learning the suite of visual features and object parts that are diagnostic of various object categories. On this account, learning what tigers tend to look like does not change children's perceptual representations of cheetahs—or how they draw them. A second possibility is that learning about new categories actually changes the similarity structure of children's visual object concepts (Goldstone, Lippa, & Shiffrin, 2001). Finally, as children learn about the hierarchical structure of object categories (i.e., living thing–animal–mammal–dog) and their typical properties (e.g., all mammals have four legs) this might differentially change which visual features take precedence in their internal representations. Future work that links children's categorization abilities with their drawing behaviors will help explore these possibilities.

This work integrates novel methods to investigate children's internal representations of object categories and how they are linked to their developing perceptual, cognitive, and motor abilities. We propose that a full understanding of how we come to produce visual abstractions will help uncover the factors that shape adult object representations.

All data and code for these analyses are available at `https://github.com/brialorelle/kiddraw`

Acknowledgements

We thank Jacqueline Quirke for help with piloting and data collection. We thank members of Stanford Language and Cognition lab. This work was funded by an NSF SPRF-FR Grant #1714726 to BLL and a Jacobs Foundation Fellowship to MCF. We also gratefully acknowledge those who made the Google QuickDraw database available.

References

Davis, A. M. (1983). Contextual sensitivity in young children's drawings. *Journal of Experimental Child Psychology*, *35*(3), 478–486.

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*.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43.

Goodenough, F. L. (1963). *Goodenough-harris drawing test*. Harcourt Brace Jovanovich New York.

Juttner, M., Wakui, E., Petters, D., & Davidoff, J. (2016). Developmental commonalities between object and face recognition in adolescence. *Frontiers in Psychology*, *7*.

Kosslyn, S. M., Heldmeyer, K. H., & Locklear, E. P. (1977). Children's drawings as data about internal representations. *Journal of Experimental Child Psychology*, *23*(2), 191–211.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis–connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*.

Kubilius, J., Bracci, S., & Op de Beeck, H. P. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS Computational Biology*, *12*(4), e1004896.

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.

Mash, C. (2006). Multidimensional shape similarity in the development of visual object classification. *Journal of Experimental Child Psychology*, *95*(2), 128–152.

Minsky, M., & Papert, S. A. (1972). Artificial intelligence progress report.

Nishimura, M., Scherf, S., & Behrmann, M. (2009). Development of object recognition in humans. *F1000 Biology Reports*, *1*.

Pike, A. W., Hoffmann, D. L., Garcia-Diez, M., Pettitt, P. B., Alcolea, J., De Balbin, R., ... others. (2012). U-series dating of paleolithic art in 11 caves in spain. *Science*, *336*(6087), 1409–1413.

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

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.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*(1), 13–19.

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

Yamins, D., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624.