# Harnessing Human Perception and Machine Vision to Examine Children's and Adult's Attempts to Draw a Cube


Clint A. Jensen[1]

Karl S. Rosengren[2]

Timothy T. Rogers[1]

Brittany G. Travers[3]

Heather L. Kirkorian[4]

[1]Department of Psychology, University of Wisconsin-Madison, United States, [2]Department of Brain and Cognitive Sciences, School of Arts and Sciences, University of Rochester, United States, [34]Occupational Therapy Program in the Department of Kinesiology and the Waisman Center, University of Wisconsin–Madison, United States, [4]Department of Human Development & Family Studies, School of Human Ecology, University of Wisconsin-Madison, United States

**Abstract**

Capturing a three-dimensional object in drawing on a two-dimensional surface is a challenging task for both children and many adults. Prior research has investigated several factors that influence this ability, including age, gender, and various spatial abilities. The current study investigates whether contemporary data science tools, including deep neural network models of vision and crowd-based similarity ratings, capture systematic latent structure in children's and adults' drawings of cubes. We then consider whether the novel metrics predict characteristics of the drawer including both the previously studied demographics age and gender, as well as children's motor skill as measured by performance on paper-folding task. Further, we examined whether predictive relationships differed as a function of the drawing medium (marker on paper, stylus on a tablet, finger on a tablet). Results varied across the three drawing tasks, with drawing with marker and paper resulting in highest predictions of motor function from the measures (59% of variance for offset, 62% of the variance for angle, *p*s < .01). Interestingly, girls overall scored better on the paper folding metrics, but higher-quality drawing scores for boys predicted better paper folding in boys, whereas the reverse was true for girls. These and other results will be discussed in terms of the potential value of this novel approach for using drawings to characterize cognitive, behavioral, and motor characteristics of children.

## Summary

Children's drawings have been viewed as a window into how children represent the spatial world (Piaget & Inhelder, 1967). One notable challenge involves accurately capturing a three-dimensional object in drawing on a two-dimensional surface, a task that is difficult for young children and many adults. Participants in such studies produce a wide variety of solutions to this task that include producing a shape that resembles a single square, a number of conjoined shapes depicting two or more sides of the cube, and even representations that appear as if the cube were unfolded as a two-dimensional figure (e.g., Cox, 1986; Lange-Küttner & Ebersbach, 2013). Given this level of disparity between drawn representations, there has been substantial research investigating the factors that influence the ability to effectively capture a three-dimensional cube and other complex objects in drawings in two dimensional drawings (Bremner et al., 2000; Cox, 1986; Kosslyn et al., 1977; Lange-Küttner & Ebersbach, 2013). These factors include age, gender, and various spatial abilities (e.g., mental rotation). In the current study we investigate the efficacy of using contemporary data science tools, including deep neural network models of vision and crowd-based similarity ratings for the assessment of children's and adults' drawings of cubes. The overall goal of our research is to examine whether our novel metrics can reveal latent structures that can enhance our understanding of children's cognitive, perceptual, and motor competence. In the present study we explore whether three new metrics obtained from innovations in machine vision and crowd-sourced human similarity judgements are predictive of age, gender, and performance on a paper-folding task. We also examined whether predictions varied as a function of the drawing medium (marker on paper, stylus on a tablet, finger on a tablet).

## Methods

**Participants**. The participants included 129 children between the ages of 2 and 9, *M*age = 4 years, 4 months, 53% female) and 25 adults (*Mag*e = 20, 76% female).

**Method**. As part of a larger project investigating children's drawing across media (Kirkorian et al. 2020) children and adults were asked to complete a series of drawing tasks (e.g., shape copying, human figure drawing, drawing of a three-dimensional translucent cube) as well as a number of perceptual motor tasks (e.g., finger and grip strength, paper folding). In this investigation we focus on the cube drawings and paper folding results. Information about the application of our novel methods to examine human figure drawings can be found in Jensen et al., under review). In addition, participants completed cube drawings using a marker on paper, using a stylus on a tablet, and using a finger on the tablet. Participants were also asked to complete three different folds on a standard piece of paper (corner fold, width fold, and length fold). For this preliminary analysis we focus on the corner fold as this one the first fold to be completed and the fold for which we have the most complete data. Folds were identified by a dotted line printed on the paper and an experimenter described and demonstrated one fold prior to the participant completing the task. Folds were assessed using a standard approach (Travers et al., 2018) that assessed the variation from the dotted line at the two edges of the folds. We used two metrics in our analyses: *offset*, mean of the absolute value of the two measures, and *angle*, absolute value of the difference between the two measures.

**Novel Metrics**. A more detailed version of our approach can be found in Jensen et al (under review). Briefly, we used two techniques to derive three novel metrics. The *first approach* uses a neural network model (VGG-19; Simonyan & Zisserman, 2014) that takes bitmap images of the cube drawings as input to extract visual features of the cube drawings. The procedure results in a re-representation of each drawing as a *machine-derived latent feature vector* that captures similarities across the neural network models' internal representations. These vectors can then be used in standard regression models to predict characteristics of the participant across different tasks. The *second approach* uses crowd-based judgements from non-experts to quickly and reliably estimate the perceptual similarity and quality of drawings. The motivation for this approach stems from the finding that human perception of drawings may be sensitive to characteristics of the drawings not captured by machine-vision approaches (Jensen et al., under review). We used two tasks, *a triadic judgment task*, where over many trials, many adult raters judge which of two drawings is most perceptually similar to a third target drawing. From such judgements, the drawings are embedded within a low-dimensional space that expresses the perceived similarity/dissimilarity amongst the drawings. Coordinates in the embedding space effectively re-represent the cube drawings as *human-derived latent feature vectors* that can again be used in regression models to predict participant characteristics across tasks. Our second human-derived perceptual judgement task presented human raters with two of the cube drawings and asked them to choose "which is the better drawing of a cube". Over many judgements of random pairs we computed the proportion of trials for which a particular cube drawing was chosen to be the 'best'. From this we derived a *quality-rank score* that represents non-expert human judgements about the quality of each cube drawing.

**Results**. Using participant age and gender as both outcome measures and control variables depending on the model, we included machine-derived vector information, human derived vector information, and the quality rank score to predict the paper folding measures using step-wise regression with all the variables and all interactions allowing the forward- and backward-selection algorithms to decide which terms to keep in the model. Results varied across the three drawing tasks, with drawing with marker and paper resulting in highest predictions from the measures (59% of variance for offset, 62% of the variance for angle, $p$s < .01). Interestingly, girls overall scored better on the folder metrics, but higher-quality scores for boys predicted better paper folding in boys, the reverse was true for girls. These and other results will be discussed in terms of the potential value of this novel approach.