

Semantic structure in communicative drawings

Kushin Mukherjee

Department of Cognitive Science
Vassar College

kumukherjee@vassar.edu

Robert X. D. Hawkins

Department of Psychology
Stanford University

rxdh@stanford.edu

Judith E. Fan

Department of Psychology
Stanford University

jefan@stanford.edu

Abstract

[jefan: Placeholder abstract: Sets this paper up as being about visual communication, that we use semantic segmentation data to investigate.]

Drawing is a versatile tool for communication, spanning detailed renderings and simple sketches. Even the same object can be drawn in different ways, depending on the context. How do people decide how to draw in order to be understood? Here we investigate the semantic structure of drawings as a window into how people deploy both perceptual information and conceptual knowledge to produce communicatively effective drawings in context. We analyzed a dataset containing drawings of real-world objects that were produced in different semantic contexts, and contained both detailed and simpler sketches of each object. We explored the hypothesis that during visual communication, people spontaneously decompose visual objects into semantically meaningful parts (e.g., chairs consist of legs, seat, and back), resulting in a tight correspondence between the organization of this semantic part knowledge and the procedure people use to sketch an object. For example, if someone aims to produce a recognizable sketch of a chair, they produce strokes that represent individually meaningful parts, e.g., seat, armrest, legs. To investigate this, we developed a web-based platform to collect dense semantic annotations of the stroke elements in each drawing. We found that: (1) people are highly consistent in how they interpret what individual strokes mean; (2) single strokes tend to represent a single part category (e.g., leg vs. leg + seat), while multiple strokes may be combined to represent an entire part category (e.g., all the legs on a chair); and (3) strokes representing the same part tend to be clustered in time, suggesting that people tend to start and finish drawing one part of an object before moving onto the next.

Keywords: sketching; cognitive science; perception

Introduction

This is where our introduction will go.

Methods

Dataset

We required sketches of the common objects created under different contexts. So we obtained sketch data from a two-player 'Pictionary'-style reference game experiment. In this experiment, a 'sketcher' aimed to produce sketches of target objects to distinguish them from three distractor objects. A 'viewer' had to guess which of the 4 images the sketch represented. The targets and distractors were chosen from a set of 32 real-world objects belonging to 4 basic-level categories: cars, chairs, dogs, and birds. Each category had 8 distinct exemplars. There were 2 main context conditions in this experiment - close and far. In the close condition, the target image and the distractors belonged to the same basic-level category. In the far condition, the target and each of the distractors belonged to a different basic-level category.

We obtained 1198 sketches for the annotation task. These sketches were represented as scalable vector graphics (SVG) images. The strokes that participants made on the canvas when creating the sketch can be represented as a concatenated string of cubic Bezier curves. Thus, the final sketch can be represented by a list of such concatenated strings, each of which corresponds to an event of the participant placing their drawing instrument on the canvas, making some marks on the canvas, and lifting the instrument off of the canvas. We were interested in collecting fine-grained annotations of these strokes, so we split strokes into sub-stroke elements, which we called splines. A single spline was equivalent to a single cubic Bezier curve, i.e., a Bezier curve with two fixed end points and two control points to control curvature. We had participants in our annotation task label each sketch's constituent splines.

Participants

We recruited a total of 326 participants via Amazon Mechanical Turk (AMT). For this experiment, participants provided informed consent in accordance with the Stanford University IRB. Participants were paid a base amount of \$0.35 and were given an additional bonus of \$0.002 for every stroke they annotated. In addition to this, they were given a \$0.02 bonus for every sketch for which they labeled all strokes.

Annotation Procedure

To collect fine-grained annotations of our sketches, we implemented a web-based Javascript annotation tool. Each participant annotated 10 sketches. We provided participants with a sketch to be annotated on a canvas as well as a category-specific menu of labels, which they were encouraged to use for the annotation task. We also provided them with the option of entering their own labels through a free-response box. The original set of images the sketcher had to discriminate between were shown to help the annotator better understand the contents of the sketch. Labeling was done by clicking on individual splines or clicking and dragging across multiple splines to highlight them before assigning them a label. Participants were encouraged to conduct their labeling of strokes in bouts — they were to highlight all the strokes corresponding to a single instance of a part before selecting a label from the menu. Participants could do the task at their own pace and continue to a subsequent sketch whenever they felt they were ready. They could choose to continue to the next trial without labeling every stroke in a sketch, but they would lose out on the completion bonus as well as the amount they would

have earned for labeling the remaining strokes. In total, we collected 3608 annotations

Analysis

Results

Inter-annotator reliability

How much agreement is there of what a spline represents between different annotators? That is, do different people see the same parts in abstract-ish sketches?

Relationship between strokes and parts

Was there a need for us to collect data at the spline-level, or is it so that the individual strokes people make correspond to singular parts? Do people need to make multiple strokes to represent a single part/instance of a part?

Stroke sequence organization

Is there any semantically meaningful manner in which people organize their strokes when sketching? Do people draw all the stroke belonging to the same part in a 'streak'? Is there spatial contiguity in how people structure their sketches?

Discussion

Acknowledgments

Tables

Figures

References