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 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 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

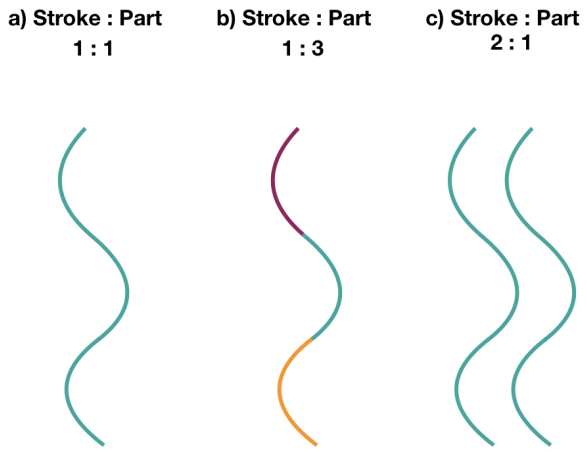


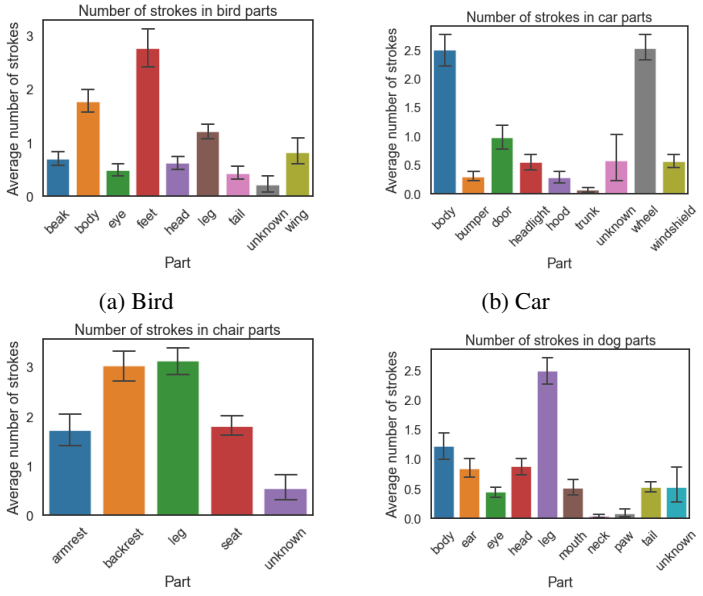
Figure 1: The three part-stroke relationships we explored. Each squiggle represents a hypothetical stroke. Different colors indicate different part-labels.

have earned for labeling the remaining strokes. In total, we collected 3608 annotations of 1195 unique sketches.

Preprocessing

After collecting annotations, we filtered out any sketches that didn't have all of its constituent splines labeled. This left us with 3319 annotations of 1190 unique sketches. Since there was some variability in the number of times each sketch in our dataset was annotated, we selected those sketches that had been annotated exactly 3 times. This left us with 764 unique sketches, each of which had been annotated 3 times. We also created unique dictionaries for each object category that mapped participant-generated labels to the most frequently occurring labels in our dataset. This helped reduce the total number of unique labels in our dataset from 228 to 24.

Given that our goal was to created an annotated dataset of sketches created under different contexts, we required that the annotations we collected through our interface be reliable. In order to assess this reliability, we looked at whether different annotators saw the same parts in these abstract sketches of objects. Specifically, we looked at inter-annotator reliability in spline labels between participants for each spline in our dataset. Reliability was measured in terms of 'agreement' on spline labels. For example, a 3/3 agreement score for a given spline meant that each of the 3 annotators applied the same label to that spline. We found that 67.85% of splines in our dataset had 3/3 inter-annotator agreement, 27.77% of splines had 2/3 agreement, and 4.38% of splines had no agreement, which means that each participant applied a different label for each of those splines. For the purposes of analysis, we set the modal label for each spline as its true label.



(a) Bird (b) Car

(c) Chair (d) Dog

Figure 2: Average number of strokes used to draw each part per category. While participants use multiple strokes to represent some parts, other parts are sometimes expressed using single strokes. It is possible that

Results

Relationship between strokes and parts

People's hierarchical organization of visual concepts, such object category membership being determined by its constituent parts, allows for robust recognition of objects in the real world. We were interested in whether people might employ similar abstractions in producing sketches of such objects as well. Since an individual stroke correspond to a person's decision to make a mark on the canvas, we looked at the relationship between strokes and the parts that they represented in our dataset. We explored 3 possible stroke-to-part relationships: a) Singular strokes correspond to singular parts, b) Singular strokes are used to convey multiple parts, that is, strokes cross semantic boundaries, and c) Multiple strokes are required to convey a single part.

We compared a) and b) by looking at within-stroke label agreement for spline labels for all strokes in our dataset. High agreement among all the splines in a given stroke would be indicative of that being used to represent a single part. On the other hand, low agreement would indicate that stroke crosses semantic boundaries and is used to represent different parts. We found that splines contained in 76.85% of strokes in our dataset shared a single label, 12.75% of strokes contained 2 labels, and less than 11% of strokes contained 3 or more labels. People, in general, tended to use their strokes to draw a single part while only sometimes utilizing a single stroke to represent multiple parts.

We also compared a) and c) by looking at the average number of strokes used to represent specific parts within a given

category of sketches. A high average number of strokes for a given part would indicate that multiple strokes are utilized to draw that part, whereas a low average would indicate that a single or few strokes might suffice in depicting that part. Figure 2 shows these part-specific stroke averages by object category. [kushin: I feel the above summary, including the figure caption for figure 2 is a little inconclusive. Thoughts on how to remedy this?]

Stroke sequence organization

[kushin: Will update this with newer z-score based analysis]

Modulation between communicative contexts

[jefan: where we would report analysis of the sketch part features (num strokes, arc length) e.g., when the far sketches are more abstract, how does that manifest in this feature representation? like, are they more similar to each other, more like "bird" and lacking object-specific details? a way of measuring this is that the centroid (euclidean norm, magnitude of the vector) is closer to the origin for far vs. close, and also that the RMSD to centroid of far sketches is smaller than for close sketches....]

Discussion

Acknowledgments

Tables

Figures

References