Communicating semantic part information in 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
UC San Diego
jefan@ucsd.edu

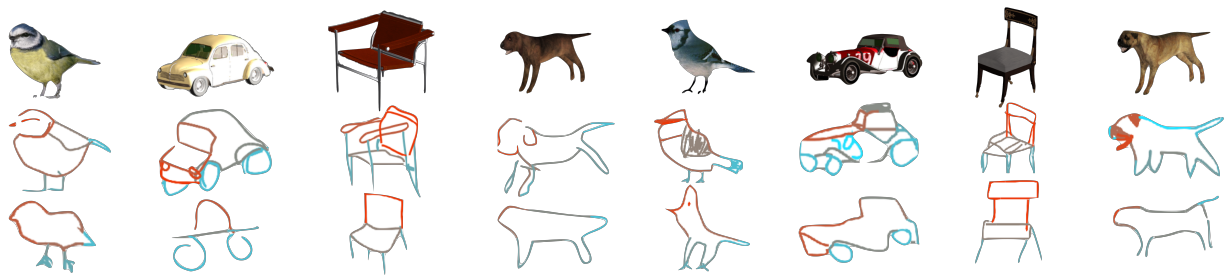

Figure 1: Objects used in communication game with example drawings below, where stroke color indicates different parts.

Abstract

We effortlessly grasp the correspondence between a drawing of an object and that physical object in the world, even when the drawing is far from realistic. How are visual object concepts organized such that we can both recognize these abstract correspondences and also flexibly exploit them when communicating them to others in a drawing? Here we consider the notion that the compositional nature of object concepts enables us to readily decompose both objects and drawings of objects into a common set of semantically meaningful parts. To investigate this, we collected data on the part information expressed in drawings by having participants densely annotate drawings of real-world objects. Our dataset contained both detailed and sparser drawings produced in different communicative contexts. We found that: (1) people are consistent in what they interpret individual strokes to represent; (2) single strokes tend to correspond to single parts, with strokes representing the same part often being clustered in time; and (3) both sparse and detailed drawings of the same object emphasize similar part information, although detailed drawings of different objects are more distinct from one another than sparse drawings. Taken together, our results support the notion that people flexibly deploy their abstract understanding of the compositional part structure of objects to communicate relevant information about them in context. More broadly, they highlight the importance of structured knowledge for understanding how pictorial representations convey meaning.

Keywords: compositionality; objects and categories; perceptual organization; sketch understanding; visual communication

Introduction

When we open our eyes, we do not experience a meaningless array of photons — instead, we parse the world into people, objects, and their relationships. The ability to represent semantically meaningful structure in our environment is a core aspect of human visual perception and cognition (Navon, 1977). As a testament to this ability, we effortlessly grasp the correspondence between a physical object in the world and a simple line drawing of it, even though such drawings lack much of the rich visual information present in real-world objects, including color and texture. How are visual object concepts organized such that they can robustly encode such abstract correspondences? Here we explore the notion that perceiving these correspondences is supported by our ability to decompose both objects and drawings into a common set of semantically meaningful parts (Biederman & Ju, 1988).

Recent advances in computational neuroscience have provided an unprecedentedly clear view into the algorithms used by the brain to extract semantic information from raw visual inputs, including drawings, exemplified by modern deep learning approaches (Fan, Yamins, & Turk-Browne, 2018; Yamins et al., 2014). Nevertheless, a major gap remains in adapting such deep learning models to emulate the structure and flexibility of human semantic knowledge (Lake, Ullman, Tenenbaum, & Gershman, 2017). A promising approach to closing this gap may be to exploit the parsimony and interpretability of structured representations that reflect how visual concepts are organized in the mind (Battaglia et al., 2018).

However, pursuit of this strategy relies upon a thorough empirical understanding of this conceptual organization and how people express this knowledge in natural behavior. We aim to contribute to this understanding by probing the expression of visual semantic knowledge in a naturalistic setting that exposes both its structure and flexibility: visual communication via drawing. This approach departs from the conventional strategy for inferring the organization of visual object concepts, which entails eliciting judgments with respect to a small number of experimenter-defined dimensions. Instead, drawing tasks permit participants to include any elements they consider relevant and combine these elements freely, yielding high-dimensional information about how people organize and deploy visual semantic knowledge under a naturalistic task objective.

Recent computational work using drawing tasks to probe visual concepts have focused on either recognition (Eitz, Hays, & Alexa, 2012; Yu et al., 2017) or generation (Ha & Eck, 2017; M. Li, Lin, Mech, Yumer, & Ramanan, 2019) of *entire* drawings. However, the question of how semantic information *within* drawings is organized has not been inves-

tigated as thoroughly (cf. L. Li, Fu, & Tai, 2018; Schneider & Tuytelaars, 2016). The goal of this paper is to present a systematic approach to analyzing the correspondence between semantic knowledge about the internal part structure of objects and the procedure by which people robustly convey this knowledge in their drawings. Specifically, this paper advances recent work investigating how drawings convey semantic information in three ways: *first*, we collect dense part annotations on freehand drawings of real-world objects, allowing an explicit focus on compositional part structure, *second*, we explore the link between this semantic structure and the dynamics of drawing production, and *third*, we examine differences in how visual semantic knowledge is expressed between contexts.

Methods

We developed a web-based crowdsourcing tool, built with `jsPsych.js` (de Leeuw, 2015), to collect dense semantic annotations of the stroke elements in drawings of real-world objects (Fig. 1).

Communicative drawing dataset

We first obtained 1195 drawings of 32 real-world objects from a previously collected experimental dataset in which pairs of participants played a drawing-based reference game (Fan, Hawkins, Wu, & Goodman, 2019).[1] Object stimuli were photorealistic 3D renderings belonging to one of four basic-level categories (i.e., bird, car, chair, dog), each of which contained eight exemplars. On each trial of the experiment, participants were presented with a shared context containing four of these objects. One participant (the sketcher) was privately cued to draw a target object so that the other participant (the viewer) could pick it out from the set of distractors. Across trials, the similarity of the distractors to the target was manipulated, yielding two types of communicative contexts: *close contexts*, in which all four objects belonged to the same basic-level category, and *far contexts*, in which objects belonged to different basic-level categories. This context manipulation led sketchers to produce relatively simpler drawings containing fewer strokes and less ink on far trials than on close trials, while still achieving high recognition accuracy in both contexts.

Prior works analyzing the semantic properties of drawing data have used a raster image representation (e.g., `*.png`), an expedient format for applying modern convolutional neural network architectures (Fan et al., 2018; Sangkloy et al., 2016; Yu et al., 2017). However, to investigate how semantic structure manifests during drawing production, it was critical to encode each drawing using a vector image format that preserves the inherently sequential and contour-based nature of drawing production (e.g., `*.svg`). Thus, each drawing in our dataset is represented as a sequence of individual strokes. A stroke is defined as the mark left by a virtual pen on a digital drawing canvas between being 'placed onto' the canvas and being 'lifted up'. We parameterized each stroke by a sequence of cubic Bézier curves, called *splines*. This format provides a compact representation of drawing data, which also preserves the sequence in which each element was produced.

[1] All materials and data are available at `https://github.com/cogtoolslab/semantic_parts`.

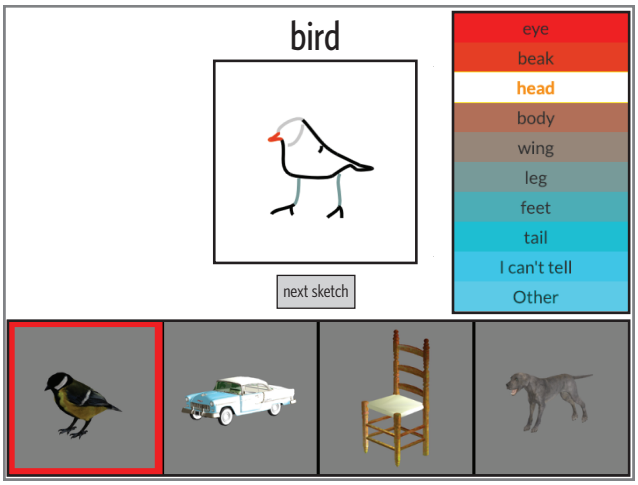

Figure 2: Annotation interface. Participants selected sub-stroke elements (splines) and tagged them with part labels.

Semantic part annotation

We crowdsourced dense semantic annotations for every spline in every stroke of the drawings from this dataset. We refer to our annotation data as *dense* because labels were provided for splines, which are at a finer level of granularity than strokes.

Participants 326 participants were recruited via Amazon Mechanical Turk (AMT) and provided informed consent in accordance with the Stanford IRB. Participants were given a base compensation of $0.35, plus $0.002 for every spline they annotated and $0.02 for every drawing they annotated completely.

Task procedure Each participant was presented with a sequence of 10 drawings that were randomly sampled from the communicative drawing dataset (Fig. 2). Their goal was to tag each spline with a label corresponding to the part it represented (e.g., seat, leg, back for a chair). To facilitate consistent tagging, participants were provided with a menu of common part labels that were associated with each basic-level category (Table 1). Participants could also generate their own part label if they believed none of the common labels applied. If any spline was too short for annotators to feasibly annotate it with their mouse cursor, it was concatenated with its neighboring splines until the resulting spline was long enough to easily select. To give participants full information about the original communicative context, we showed the drawing with the same array of four objects that the original sketcher had viewed, with the target object highlighted in red.

Data preprocessing We first standardized all 304 distinct labels provided by participants, mapping them to a common set of 24 part labels that applied to all objects in the dataset. This common set was defined as the superset of all labels that appeared in the part menu in the annotation task. Although most labels provided already exactly matched one in the common set (i.e., 90.1%), participants were permitted to assign their own custom label, resulting in additional lexical variation that we collapse over in the current analysis. For example, some custom labels were either synonymous with or more specific than one of the common labels (e.g., 'leg support', 'foot', or 'strut' for 'leg'). We manually constructed a part dictionary to map such custom labels to one of the common ones, ensuring a consistent level of granularity for all spline labels. We only examined drawings that were annotated by at least three distinct participants, providing a consistent way to evaluate annotation consistency across splines. To reduce bias due to missing data, we also restricted our analyses to annotation trials in which the drawing was completely annotated (i.e., all splines were tagged). After applying all preprocessing, our resulting dataset consisted of 864 drawings that had been completely annotated 3 times.

Results

How well do viewers agree on what strokes mean?

Before proceeding to use these annotations to examine how semantic information is conveyed during drawing production, we conducted a basic check of inter-annotator consistency. Specifically, we examined how often different annotators agreed on what each spline in a drawing represented. We found that 95.6% of all splines received the same label by at least two of the three annotators, and 67.8% of all splines received the same label by all three annotators. This shows that the way viewers interpret which part each stroke represents is systematic, validating our general approach. Further, it suggests that sketchers may exploit this systematicity to produce strokes that they expect viewers to interpret consistently. In subsequent analyses, we collapsed over inter-annotator variation: we assigned the modal label to splines to which at least two annotators had given the same label; for the remaining 4.4% of splines, we sampled one of the three labels provided.

How do strokes correspond to parts of objects?

When composing a recognizable drawing of a real-world object, how do people decide what information to convey with each stroke? A natural possibility is that their actions closely correspond to the part structure of that object. Concretely, we hypothesized that most strokes in our dataset would *not* cross part boundaries: that all splines within a given stroke would be assigned the same part label. Conversely, because depictions of parts can be arbitrarily detailed, and some parts re-occur throughout an object (e.g., multiple legs on a bird, chair, or dog), we hypothesized that there would often be more than one stroke per part (Fig. 3A).

category	part labels
bird	eye, beak, head, body, wing, leg, feet, tail
car	bumper, headlight, hood, windshield, window, body, door, trunk, wheel
chair	backrest, armrest, seat, leg
dog	eye, mouth, ear, head, neck, body, leg, paw, tail

Table 1: Part labels provided to annotators.

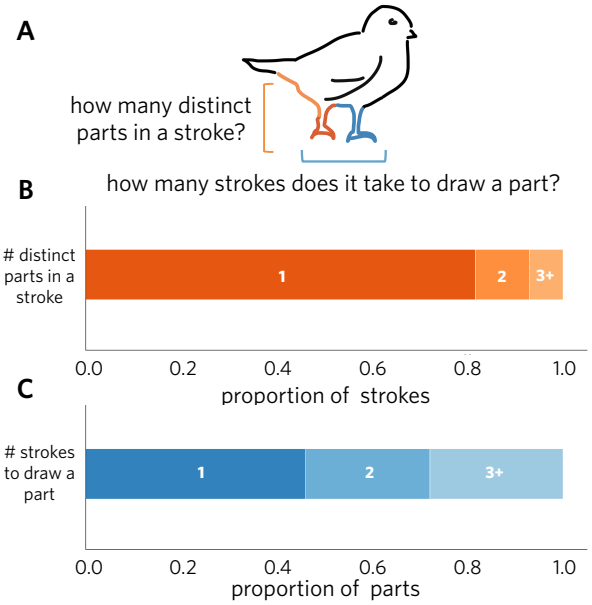

Figure 3: (A) Analyzing the correspondence between strokes and part labels: number of unique part labels assigned to different splines within the same stroke and number of different strokes used to draw each part. (B) Distribution over number of part labels within a stroke. (C) Distribution over number of strokes used to draw a part.

To evaluate the first hypothesis, we computed the number of unique part labels across all splines within each stroke. We found that for 81.6% of the strokes in our dataset there was only one part label; the remaining 18.4% of strokes were associated with two or more labels (Fig. 3B). In other words, most strokes represented exactly one part, but in a minority of cases they spanned multiple parts (e.g., a single stroke connecting the head and body of a bird, or an armrest and leg of a chair). We were concerned, however, that these proportions were inflated by strokes with very few splines.[2] To address this concern, we constructed a null model controlling for the number of splines. Part labels were randomly sampled from the full list of parts in the drawing such that each spline was equally likely to represent any part regardless of the stroke it belonged to. In simulations from this null model, only 55% of strokes corresponded to a unique part while 45% of strokes spanned multiple parts. Thus, individual strokes in our dataset were much more likely to correspond to a single part (i.e., not cross part boundaries)

[2]The modal number of splines per stroke (20% of cases) was 1, but there was a long tail; the mean number was 2.6.

than would be expected under random assignment of part labels to splines.

To evaluate the second hypothesis, we computed the number of strokes that were used to represent each part of an object (Fig. 3C). We found that 46.1% of parts were depicted using exactly one stroke, 26.0% using exactly two strokes, 11.3% using exactly three strokes, and 16.6% using four or more strokes. Thus, nearly half the time, a single action was sufficient to depict an entire object part. However, the remaining 53.9% of the time, more than one stroke was required to depict an entire part, which would be expected for those parts that consisted of multiple disconnected subparts within an object (e.g., wheels of a car, paws of a dog).

The findings so far show that the information people convey with each stroke systematically corresponds to the parts that objects contain. We next sought to understand how these properties may vary between drawings generated in different communicative contexts. Indeed, strokes spanning multiple parts were slightly more common in drawings produced in far contexts (19.4%, CI: [17.9%, 20.9%]) than close contexts (17.6%, CI: [16.1%, 18.8%][3], p= 0.07), suggesting that sketchers were somewhat more likely to use a single stroke to represent multiple contiguous parts in a context where a sparser drawing would be sufficient. And the proportion of parts requiring more than one stroke was slightly higher for close drawings (55.8%, CI: [53.7%, 58.6%]) than far drawings (52.0%, CI: [49.9%, 54.6%],p = 0.02), suggesting that sketchers may have included more detail per part in close drawings to distinguish the target object from similar distractors.

Do strokes representing the same part tend to be produced in succession?

In the previous section we discovered that slightly more than half of the parts in our dataset were depicted using multiple strokes. This result raised the question: to what extent are strokes depicting the same part drawn in succession, or interleaved among strokes depicting other parts?

To investigate this question, we estimated the mean length of 'streaks' containing strokes depicting the same part. First, we collapsed across the spline annotations examined in the previous section and represented each stroke by the modal part label assigned to its splines. We represented each drawing as the sequence of these part labels, and defined *part streak length* to be the number of consecutive strokes annotated with the same part label.[4] For example, in the drawing shown in Fig. 4A, two 'leg' strokes were placed before moving on to the 'foot', giving a streak of length 2. Finally, we averaged these streak length values over every drawing in the dataset to obtain our statistic.

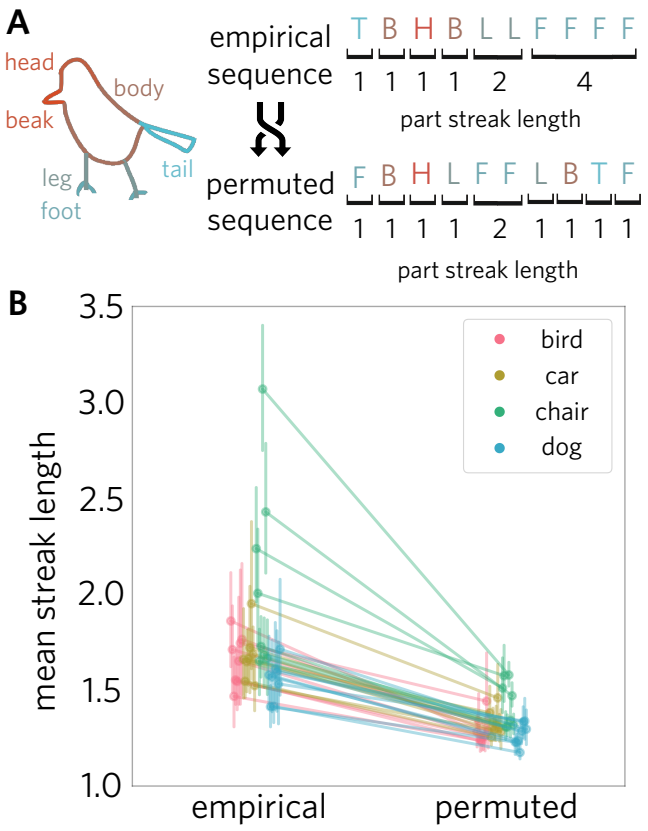

Figure 4: (A) Analysis of sequence in which strokes depicting each part were drawn. (B) Comparison of mean length of streaks consisting of strokes that depict the same part with null distribution of permuted stroke sequences.

To evaluate whether the empirical part sequences were more structured than expected if parts were drawn at random, we constructed a null model to serve as a baseline. For this null model, we permuted the part sequence such that the number of instances of each part was preserved, but the temporal structure was disrupted (Fig. 4A). We generated a null distribution of streak lengths for each drawing by repeating this permutation procedure 1000 times and measuring the mean streak length for each permutation. Finally, we obtained a z-score for each drawing by computing where the empirical streak length fell in the permuted streak length distribution. A drawing with a z-score near 0 had a streak length that was commonly obtained by placing strokes in a random order, while a drawing with a higher z-score is more structured than expected under the null.

We found that the empirical streak length was reliably higher for all objects than that of the permuted sequences (mean z-score across drawings: 2.07, CI: [1.90, 2.23]; Fig. 4B), and higher for the close drawings (mean z-score: 2.58; CI: [2.26, 2.90]) than far drawings (mean z-score: 1.56; CI: [1.38, 1.74]). The lower streak length for far drawings is consistent with their lower stroke count overall—when only

[3] 95% confidence intervals were estimated via stratified bootstrap resampling (N=1000 iterations) of drawings within each context condition.

[4] We excluded 78 out of the 864 drawings where this measure was not well-defined, i.e. sketches containing only one stroke or part label, or containing fewer than two strokes sharing the same part label.

one or two strokes are used per part, there is a ceiling on the mean streak length. However, when sketchers do use multiple strokes to convey a single part (i.e., because there are multiple subparts, or to add more detail), they tend to draw these in succession before moving on to a different part. These results suggest more broadly that the procedure by which people convey semantic information in drawings is organized by the part structure within objects.

How is part information emphasized in different communicative contexts?

Our findings so far bear on how the way people compose communicative drawings of objects reflects their semantic knowledge of the parts those objects are composed of. A key consequence of such semantically organized part knowledge is that it naturally supports flexible expression across different communicative contexts. For example, when communicating about a chair in a far context containing objects from other basic-level categories, sketchers may include only the essential information to indicate the presence of certain parts (e.g., armrests) that distinguish it at the category level. On the other hand, when communicating about that same chair in a close context containing other, perceptually similar, chairs sketchers may emphasize aspects of parts that distinguish it at the object level (e.g., the curvature of the armrests), by applying more strokes and/or more ink in each stroke.

We hypothesized that sketchers emphasize part information to preserve relevant distinctions in context. To explore this possibility, we asked the following questions: (1) How similarly is object-specific part information emphasized in both close and far contexts? (2) How do differences in how part information is emphasized *between* contexts affect how discriminable those drawings are?

To investigate these questions, we represented each drawing by a 48-dimensional *part-feature vector* that contained information about: (a) how many strokes and (b) how much total ink was allocated to each of the 24 unique part labels in our dataset. Specifically, the first 24 elements of each part-feature vector contained the number of strokes allocated to each part, and the remaining 24 contained the total arc length of all strokes allocated to each part. Because our primary goal was to understand *relative* differences in how much emphasis was placed on each part across drawings in our dataset, we first z-scored the raw stroke-count and arc-length measurements within each feature dimension, thereby mapping all features to the same unit-variance scale. We then collapsed across drawings within each object-context combination, yielding 64 average part-feature vectors (i.e., 32 objects x 2 context conditions).

Similar part information emphasized across different communicative contexts In order to investigate to what extent similar object-specific part information is emphasized in different communicative contexts, we computed the matrix of Pearson correlations between part-feature vectors. Formally, this entailed computing: $R_{ij} = {}^{cov(\vec{r}_i,\vec{r}_j)}/\sqrt{var(\vec{r}_i)\cdot var(\vec{r}_j)}$, where

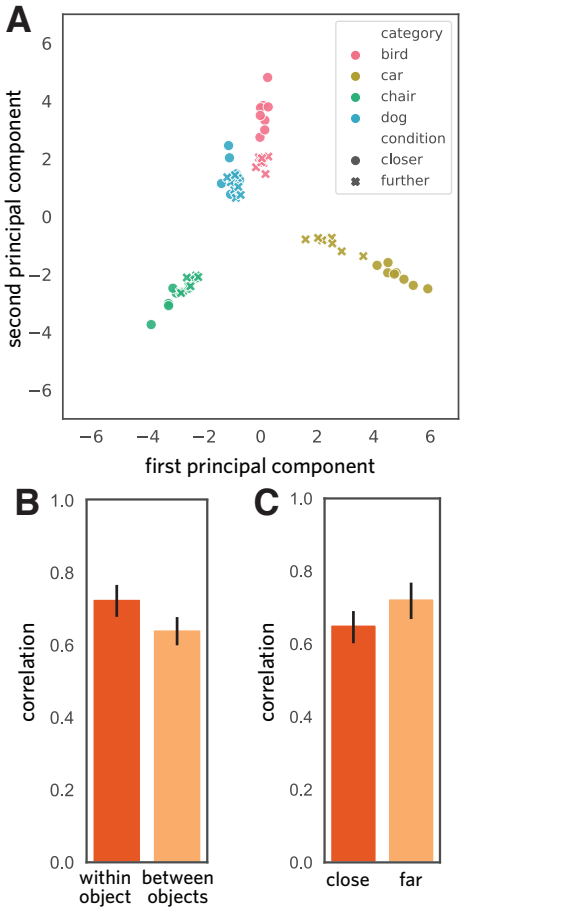

Figure 5: (A) Layout of mean part-feature vectors for each object-context combination, projected onto top two principal components. (B) Comparison of feature similarity between close and far drawings of the same object, relative to close and far drawings of different objects within a category. (C) Comparison of feature similarity between far drawings of objects within a category, relative to close drawings. Error bars reflect 95% CIs.

$\vec{r}_i$ and $\vec{r}_j$ are the mean part-feature vectors for the ith and jth object-context combinations, respectively.

While close and far drawings of an object differed in their overall amount of detail, we hypothesized that they would still emphasize part information in similar ways. Specifically, insofar as similar object-specific part information is emphasized in both close and far drawings of the same object, we predicted higher correlations between close and far part-feature vectors for the *same* object than for close and far part-feature vectors of *different* objects. Consistent with this, we found strong correlations between the feature vectors for close and far drawings of the same object ($r = 0.73$, CI: [0.68, 0.77][5]), which were significantly stronger than close and far drawings of *different* objects ($r = 0.64$, CI: [0.60, 0.68]; same objects vs. different objects: $p < 0.001$). These results show that close and far drawings of the same object exhibit similar

[5]95% confidence intervals were estimated via stratified bootstrap resampling (N=10000 iterations) of drawings within each object-context combination.

patterns of emphasis across different parts, and this similarity exceeded that expected due to merely being members of the same basic-level category (Fig. 5B).

Detailed drawings are more distinct from each other than sparser drawings While the above findings showed that close and far drawings of the same object exhibit similar patterns of emphasis on different parts, close drawings contain greater emphasis on these parts overall than far drawings (i.e., contained more and longer strokes). How were these additional strokes being spent?

We hypothesized that the additional part information provided in close drawings was being distributed across parts in different ways for different objects, thereby making them more distinguishable from one another in feature space. To evaluate this possibility, we computed the mean correlation between the part-feature vectors of close drawings of objects in a given category and compared this value with the mean correlation between far drawings of exactly the same objects. We found that close drawings were less similar to one another than far drawings were (close similarity: $r = 0.65$, CI: [0.60, 0.69]; far similarity: $r = 0.73$, CI: [0.67, 0.77]; close vs. far: $p = 0.007$), suggesting that sketchers discern which parts are most diagnostic of the target object among highly similar distractors and emphasize these parts accordingly (Fig. 5C). This was particularly apparent when we visualized the spatial layout of part-feature vectors: whereas far drawings were clustered closer together and near the origin, close drawings were spread further apart from other members of the same category and further from the origin (Fig. 5A). Observing these contextual differences is all the more remarkable given that this feature representation captures only the *amount* of emphasis allocated to each part during drawing production, setting aside their visual properties.

Discussion

In this paper, we explored how the way people compose communicative drawings of objects reflects their semantic knowledge about what objects are composed of. To accomplish this, we first collected dense semantic annotations of sub-stroke elements in communicative drawings of real-world objects that were produced in different contexts. This allowed us to interrogate the internal semantic structure within drawings, and relate this structure to the dynamics of drawing production in a naturalistic visual communication task. Overall, we found that: (1) people are highly consistent in how they interpret what individual strokes represent; (2) single strokes tend to correspond to single parts, with strokes representing the same part tending to be clustered in time; and (3) both detailed and sparse drawings of the same object emphasized similar part information, with detailed drawings of different objects tending to be more distinct from one another than simpler ones. Taken together, our results support the notion that people deploy their abstract understanding of the compositional part structure of objects in order to select actions to communicate relevant information about them in context.

These findings are resonant with classic and recent work that has argued for the importance of compositionality in human perception and cognition in general (Biederman, 1987; Battaglia et al., 2018; Lake et al., 2017), and for visual production in particular (Lake, Salakhutdinov, & Tenenbaum, 2015). However, unlike prior work which focused on the production of abstract symbols (Lake et al., 2015), we consider the challenge of how people transform perceptually grounded representations of real-world objects into procedures for producing figurative drawings that communicate not only what they see and know about them, but also what is relevant in context.

Our work is also related to recent progress in the development of computational models of drawing production (Ha & Eck, 2017; M. Li et al., 2019). While results from these efforts have been galvanizing, the development of principled metrics by which to rigorously evaluate how well they emulate human drawing behavior has not kept pace. By interrogating in detail how humans encode semantic information into their drawings, and flexibly adjust their production behavior in different contexts, this paper presents a first step towards such a set of behavioral metrics. Having such metrics is important because they would enhance our ability to distinguish between generative models, and thereby help advance further model development. It would thus be valuable to apply up our analytical approach to the large drawing datasets (Eitz et al., 2012; Sangkloy et al., 2016; Jongejan, Rowley, Kawashima, Kim, & Fox-Gieg, 2017) that have provided the basis for these modeling approaches.

In ongoing work, we are extending our analysis of how different part information is expressed in drawings beyond simple effort cost measures (i.e., number of strokes, amount of ink) to encompass content and style information (e.g., the shape of a bird's wing, caricaturization of a chair's armrest). We expect that augmenting current vision models with a combination of the requisite semantic part knowledge and the ability to discern perceptual properties of these parts, such as style, will enable us to build models that parse drawings in a more human-like way. More broadly, achieving this synthesis will lead to both more robust artificial intelligence and a deeper understanding of human cognition and behavior.

Acknowledgments

KM was supported by the Department of Cognitive Science at Vassar College through its Humanities in Cognitive Science program and the Center for the Study of Language and Information at Stanford University. RXDH was supported by the National Science Foundation Graduate Research Fellowship (DGE-114747).

All code and materials available at:
`https://github.com/cogtoolslab/semantic_parts`

References

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., ... others (2018). Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*(2), 115.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*(1), 38–64.

de Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.

Eitz, M., Hays, J., & Alexa, M. (2012). How do humans sketch objects? *ACM Trans. Graph.*, *31*(4), 44–1.

Fan, J., Hawkins, R., Wu, M., & Goodman, N. (2019). Pragmatic inference and visual abstraction enable contextual flexibility during visual communication. *arXiv preprint arXiv:1903.04448*.

Fan, J., Yamins, D., & Turk-Browne, N. (2018). Common object representations for visual production and recognition. *Cognitive Science*.

Ha, D., & Eck, D. (2017). A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477*.

Jongejan, J., Rowley, H., Kawashima, T., Kim, J., & Fox-Gieg, N. (2017). *Google Quickdraw.* Retrieved from `https://quickdraw.withgoogle.com/`

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.

Li, L., Fu, H., & Tai, C.-L. (2018). Fast sketch segmentation and labeling with deep learning. *IEEE computer graphics and applications*.

Li, M., Lin, Z., Mech, R., Yumer, E., & Ramanan, D. (2019). Photo-sketching: Inferring contour drawings from images. *arXiv preprint arXiv:1901.00542*.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383.

Sangkloy, P., Burnell, N., Ham, C., & Hays, J. (2016). The sketchy database: learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (TOG)*, *35*(4), 119.

Schneider, R. G., & Tuytelaars, T. (2016). Example-based sketch segmentation and labeling using crfs. *ACM Transactions on Graphics (TOG)*, *35*(5), 151.

Yamins, D. L., 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.

Yu, Q., Yang, Y., Liu, F., Song, Y.-Z., Xiang, T., & Hospedales, T. M. (2017). Sketch-a-net: A deep neural network that beats humans. *International Journal of Computer Vision*, *122*(3), 411–425.