Drawings of specific objects and object categories drive different visual recognition patterns

*Justin Yang[a,b], *Holly Huey[d,b], Xuanchen Lu[c,b] and Judith E. Fan[a,b]

[a]Department of Psychology, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305, USA

[b]Department of Psychology, UC San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

[c]Department of Computer Science, Cornell Tech, 2 West Loop Rd, New York, NY 10044

[d]Adobe, 601 Townsend Street San Francisco, CA 94103

Corresponding author: Judith Fan (jefan@stanford.edu)

Keywords: visual communication, concepts, abstraction, semantic memory, sketching, crowdsourcing

Word count: 6826 words

Author note: A subset of these findings were published as a conference paper in the Proceedings of the 43rd Annual Meeting of the Cognitive Science Society (Yang and Fan, 2021).

Abstract

Human conceptual knowledge about objects is not unitary, but spans multiple levels of semantic abstraction (e.g., superordinate vs. basic-level) and can be evoked via multiple routes (e.g., visual vs. linguistic cues). Here we leverage a drawing paradigm to characterize what information about a visual object concept is prioritized depending on what level of abstraction is relevant and whether this concept was activated by an image or text-based cue. First, we elicited drawings from participants who either had the goal of drawing a specific exemplar or a general object category, half of whom were cued with a color photograph and the other half of whom were cued with a category label. We found that both goal and cue type impacted how detailed the resulting drawings were. In a follow-up study, a different group of participants attempted to guess the category that each drawing was intended to convey, providing a measure of the category-diagnostic information in these drawings. A separate group of participants also attempted to match each photo-cued drawing with its corresponding photo cue, providing a measure of the exemplar-diagnostic information in these drawings. We found that label-cued category drawings were the most categorizable at the basic level, whereas photo-cued exemplar drawings were the least categorizable. However, these photo-cued exemplar drawings could be matched more often with their corresponding photo than with category drawings, suggesting that these drawings contained diagnostic information about object identity that did not necessarily support categorization. To gain further insight into what properties of these drawings might have given rise to these differences in recognizability, we obtained detailed annotations of which object part each pen stroke in each drawing represented. This stroke-level analysis found that drawings produced under different conditions emphasized the parts of objects in reliably different ways, which could partially account for differences in the recognizability of drawings across conditions. Taken together, these findings highlight novel trade-offs between different levels of semantic abstraction during visual communication.

Public significance statement: This study demonstrates that drawings of objects emphasize different visual features depending on whether people are trying to communicate about specific instance or a general category. When asked to draw a specific object, people include details that distinguish that item from others. When asked to draw a general category, they emphasize features that are more broadly shared across exemplars of that category. These different goals lead to drawings that are either better for identifying individual examples or for recognizing general categories. More generally, these findings highlight that drawings do not simply reflect the way objects look, but also what a person knows and what information they wish to communicate.

Keywords: visual communication, sketch production, sketch recognition, sketching, concepts, abstraction, semantic memory, crowdsourcing

1 Introduction

Tools for expressing ideas in visual form have been critically important throughout human history, with some of the earliest figurative art dating back to at least 35,000 years ago Aubert et al. (2014); Hoffmann et al. (2018); Tylén et al. (2020). In addition to being among the oldest techniques for producing images, drawing is also among the most versatile techniques for encoding what people perceive and know about the visual world, ranging from detailed depictions to schematic diagrams (Fan et al., 2023).

A key challenge is thus to understand the cognitive basis for this versatility, including how humans use drawings to express conceptual knowledge at different levels of abstraction (Mukherjee et al., 2023). Consider, for example, a drawing of a specific object and one corresponding to the broader category an object belongs to. In language, people often differentiate between these levels of semantic abstraction by producing different labels (e.g., "cat" rather than "animal" or "Siamese") depending on what would be informative in context (Murphy and Smith, 1982; Rosch, 1973; Degen et al., 2020). This flexibility reflects the understanding that objects possess both features that reliably predict category membership across many exemplars, as well as features that distinguish specific instances but are less reliable cues regarding category boundaries (Rosch and Mervis, 1975; Rosch et al., 1976; McClelland and Rumelhart, 1985; Logothetis and Sheinberg, 1996; McClelland and Rogers, 2003; Palmeri and Gauthier, 2004; Goldstone, 2003). When people produce drawings at different levels of semantic abstraction, do they selectively emphasize these different types of features? While it is plausible that drawings of specific objects differ from those of general categories, it remains unclear which features end up being included in one kind of drawing but not the other (Fan et al., 2020).

Existing evidence favors different possibilities. Prior work investigating categorization has found evidence that people assign weight to features differently depending on whether they are making between-category or within-category distinctions. If similar feature-weighting mechanisms operate during drawing production, drawings of exemplars might emphasize features that distinguish individuals at the expense of category-diagnostic ones (Medin and Schaffer, 1978; Nosofsky, 1986). However, other work has found that people who have expert knowledge about objects can just as easily make basic-level and subordinate-level distinctions in their domain of expertise (Tanaka and Taylor, 1991), suggesting that both more general and specific features can be jointly activated and used to drive categorization behavior. If similarly integrated and accessible representations of objects are recruited during drawing production, then drawings of exemplars might contain both features that are diagnostic at both the category level and instance level.

Moreover, conceptual knowledge about objects can be evoked via multiple routes, including both visual and linguistic cues (Potter, 1976). Prior work examining object concepts has generally relied upon linguistic cues (e.g., text labels) to probe semantic memory, often measured by having participants produce lists of attributes possessed by an object (Rosch et al., 1976; Tversky and Hemenway, 1984; Garrard et al., 2001; Murphy, 2004; McRae et al., 2005) or make speeded discrimination and categorization judgments about them (Lupyan and Thompson-Schill, 2012; Boutonnet and Lupyan, 2015). However, verbal feature-generation tasks can systematically miss important aspects of conceptual knowledge, including implicit spatial and perceptual information (Wu and Barsalou, 2009). Similarly, semantic feature production norms reveal that many conceptually important features (e.g., "has a neck" for cats vs. giraffes) are rarely listed in verbal feature-generation tasks (McRae et al., 2005; Vinson and Vigliocco, 2008; Suresh et al., 2025). By contrast, investigations of object concepts using visual cues (e.g., photographs) have focused primarily on discrimination and categorization judgments, rather than production-based methods that might more fully reveal the contents of these visually evoked representations (Rosch, 1973; Murphy and Smith, 1982; Biederman, 1987; Edelman et al., 1999; Fan et al., 2018a). Studies with semantic dementia patients using production tasks have revealed the hierarchical organization of conceptual knowledge in ways that recognition tasks could not access, with representations of more fine-grained distinctions between objects degrading before representations of coarser distinctions (Bozeat et al., 2003; Rogers et al., 2004). It is thus of theoretical importance to use both production and recognition methods to understand what information are brought to mind when people are cued with a label for an object as when they are cued with an image of it.

Here we leverage a drawing paradigm to characterize what information about a visual object concept is prioritized depending on whether a person wishes to communicate about that object at the exemplar or category levels, and whether they have recently seen an instance of that object or not. Drawing offers unique methodological advantages for studying how conceptual knowledge is activated and used. Unlike verbal methods, it more readily exposes information about spatial relations and other visual properties that someone might not be likely to mention. Unlike recognition tasks, it requires a person to actively generate and prioritize certain features. We use drawing production to distinguish two main hypotheses. The first hypothesis is that information about a specific exemplar is integrated with more general information about the category it belongs to, such that drawings meant to depict particular objects would effectively convey category membership just as well. The second hypothesis is that exemplar-specific and category-level information might be dissociable in people's drawings, such that drawings meant to depict a specific object might contain features that are diagnostic of its identity without necessarily being more categorizable. We further explore the degree to which people include information at both levels of abstraction depending on whether people have immediate visual access to an object's appearance or must rely on semantic memory alone.

To tease the above hypotheses apart, we investigated drawing production and comprehension for a diverse set of visual object categories (32 categories) and exemplars (32 exemplars per category). First, we elicited drawings from participants who either had the goal of drawing a specific exemplar or a category, half of whom were cued with a color photo and the other half of whom were cued with a category label. Next, a new group of participants attempted to guess the category that each drawing was intended to convey, providing a measure of the category-diagnostic information in these drawings. Then another group of participants attempted to match each photo-cued drawing with its corresponding photo cue, providing a measure of the exemplar-diagnostic information these drawings contained. We found that label-cued category drawings were the most categorizable at the basic level, whereas photo-cued exemplar drawings were the least categorizable. However, while these photo-cued exemplar drawings were more easily matched to their corresponding reference photo than category drawings, the details that made them more identifiable did not necessarily facilitate categorization. Finally, to understand which features drove differences in recognizability, we recruited participants to segment each drawing by labeling strokes with the object parts they appeared to represent. Taken together, our findings advance understanding of how task goals and visual access jointly shape what information people prioritize when producing drawings to communicate about visual concepts.

2 Method

We conducted the behavioral studies reported in this work in six stages, detailed below. Section 2.1 (*Sketch production*) describes the approach we used to elicit drawings under different goal and cue-type conditions. Section 2.2 (*Sketch categorization*) and Section 2.3 (*Sketch identification*) describe the procedure used to evaluate how well those drawings could evoke either the category it referred to, or the exemplar it was intended to depict. Section 2.4 (*Photo typicality evaluation*) describes our approach for obtaining typicality ratings for each image cue, which we then used to explore the degree to which the semantic information contained in photo-cued drawings varied as a function of cue prototypicality (Rosch, 1973). Section 2.5 (*Object-part label elicitation*) and Section 2.6 (*Sketch-part annotation*) describe the procedure used to characterize the semantic content in each drawing by obtaining annotations indicating what object part each drawn stroke appeared to represent.

2.1 Drawing production

2.1.1 Participants

We recruited 384 participants (128 female, 25.9 years) to participate in our study via Prolific. Each participant received \$6.00 for their participation (approx. \$12/hr). We did not exclude data from any participant, as none met our pre-registered exclusion criteria. All participants in this and subsequent studies provided informed consent in accordance with the cognizant IRB. [1]

2.1.2 Transparency and openness

All sample sizes were intended to exceed the estimated minimum sample size needed to resolve the effect of our experimental manipulations on the recognizability of drawings. These estimated minimum sample sizes were determined based on previous work using a similar stimuli and experimental procedures (Mukherjee et al., 2023; Huey et al., 2023; Fan et al., 2018b). The rationale for exceeding these minimum sample sizes was to enable exploration of the content of these drawings, especially the parts of objects participants chose to include, for which there is less clarity regarding the expected effect size of our experimental manipulation based on prior work that has explored related questions using drawing tasks (Mukherjee et al., 2019; Bainbridge et al., 2019; Long et al., 2024). All code and materials available at: `https://github.com/cogtoolslab/photodraw32`.

2.1.3 Stimuli

We designed our stimulus set to be compatible with commonly used benchmark datasets containing digital freehand drawings of real-world objects (Sangkloy et al., 2016; Eitz et al., 2012). Out of the 125 object categories in the *Sketchy* database (Sangkloy et al., 2016), we selected 32 categories spanning a wide range of familiar concepts and approximately balanced with respect to animacy (living/nonliving), size (large/small), familiarity (high/low), and naturalness (natural/artificial). These categories were: *airplane, ape, axe, blimp, bread, butterfly, car (sedan), castle, cat, cup, elephant, fish, flower, hat, hotdog, jack-o-lantern, jellyfish, kangaroo, lion, motorcycle, mushroom, piano, raccoon, ray, saw, scorpion, skyscraper, snake, squirrel, tree, windmill, and window*. Then, out of the 100 photographs from each category in the *Sketchy* database, we selected 32 images that displayed variation with respect to both category-orthogonal properties (e.g., pose, viewpoint) as well as category-relevant properties (e.g., typicality).

[1] The predefined exclusion criteria for this and subsequent experiments can be found in the GitHub repository for this project: [URL BLINDED FOR REVIEW].

2.1.4 Task Procedure

Participants used their mouse cursor to produce black-and-white drawings of different object concepts using an online drawing interface, as in prior work (Fan et al., 2018b). We manipulated cue type and goal using a 2x2 between-participants design, such that each participant was pseudo-randomly assigned to a cue-type (i.e., photo vs. text label) and abstraction-level (i.e., exemplar vs. category) condition (Fig. 2A; N=96 participants per condition). Participants in the photo-exemplar group were instructed to: "*Make a drawing that would help someone else looking only at your drawing guess which image you were prompted with out of a lineup containing other similar images.*" By contrast, participants in the photo-category group were instructed to: "*Make a drawing that is recognizable, but not one that could be matched to the image I was shown.*" Participants in the label-category group were instructed to: "*Make a drawing that would help someone else looking only at your drawing guess which word you were prompted with.*" Finally, participants in the label-exemplar group were instructed to visualize and "*Draw a* specific *object, rather than a general object category.*"

Each participant produced 32 drawings, one per category. To equate the total amount of preparation time participants in all four groups had before beginning their drawing, the cue was always presented for 8 seconds and then removed before participants could begin their drawing. The sequence in which categories appeared across trials was randomized across participants, but the number of times a given photo was presented was balanced, such that each photo served as the cue three times across all photo-cue experimental sessions. Our resulting dataset contained 12,288 drawings with 382 drawings per category (Fig. 1). At the end of the session, participants were asked to report their own level of drawing skill ("How skilled do you consider yourself to be at drawing?") using a 7-point Likert scale.

2.2 Sketch categorization

The purpose of the sketch categorization study was to measure how well each drawing evoked the object category it was intended to convey.

2.2.1 Participants

347 participants (104 female, mean age = 24.8 years) were recruited via Prolific and received \$3.00 for their participation (approx. \$12/hr). Of these participants, 15 dropped out of the study early and one participant failed to meet our predefined exclusion criteria. The data from these 16 sessions were excluded from further analysis, yielding 331 complete sessions.

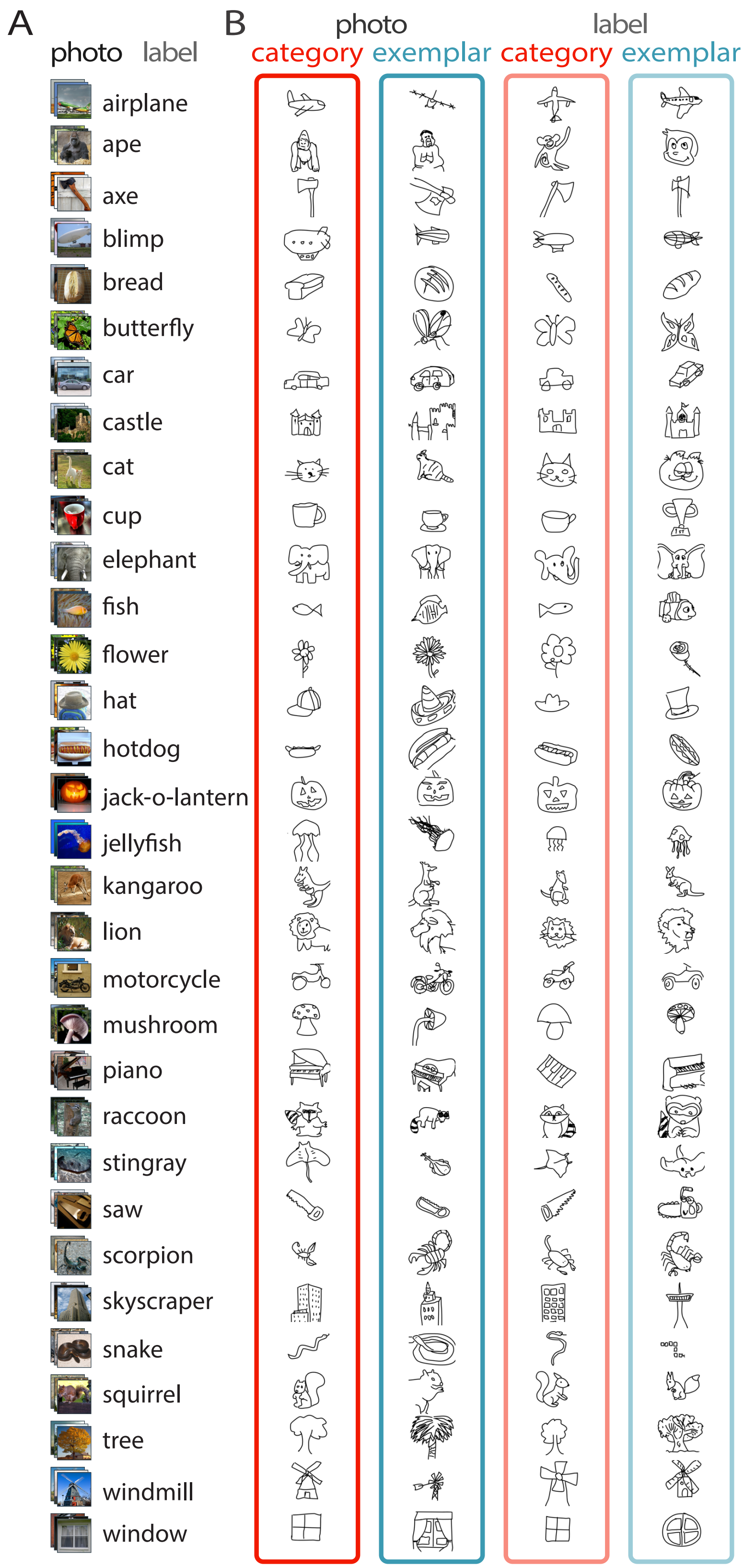

Figure 1: Example drawings produced in each condition.

2.2.2 Task procedure

On each trial, participants were presented with a randomly sampled drawing from the sketch production study and asked to select the category label that best matched it from among the full set of 32 category labels (*"Which category does this drawing belong to?"*; Fig. 3A). Each participant completed 128 trials, excluding 4 catch trials. Overall, this procedure yielded 42,368 judgments, with each drawing having received at least 3 categorization responses.

2.3 Sketch identification

The purpose of the sketch identification study was to measure how well each photo-cued drawing evoked the exemplar used to elicit it.

2.3.1 Participants

160 participants (36 female, mean age = 24.3 years) were recruited via Prolific to complete a recognition task in which they were shown a photo-cued drawing and were asked to match the drawing to one of 8 possible images. Each participant received $5.00 for their participation in the 25-minute study (approx. $12/hr). Of these participants, three dropped out of the study early and one failed to meet our preregistered exclusion criteria. The data from these four sessions were excluded from further analyses.

2.3.2 Task procedure

Each session consisted of 128 trials. On each trial, participants were presented with a randomly sampled photo-cued drawing from the sketch production study and eight color photos and were instructed to select the photo they believed the drawing was intended to depict (Fig. 3D). One of these photos was the image cue actually used to elicit the drawing and the other seven distractor images were selected from among the remaining 31 photos belonging to the same category. To make exemplar-level discrimination sufficiently challenging, we identified the seven images that were most visually similar to the original cue using an automated procedure. Specifically, we estimated the similarity between every pair of photos in our stimulus set by calculating the correlation between feature-vector representations of each image computed by a deep convolutional neural network that had been trained on a large and independent set of photographs (Simonyan and Zisserman, 2014; Deng et al., 2009). We then found the seven images whose feature vectors had the highest correlation (i.e., were the nearest neighbors) with that of the actual cue. We additionally included four catch trials containing highly identifiable drawings that we expected any sufficiently engaged participant to be able to succeed on. These four catch trials were interleaved among the other drawings and were the same for all participants.

2.4 Photo typicality evaluation

The purpose of the typicality evaluation study was to measure the relative prototypicality of the exemplars contained in the image cues used in the current study (Rosch, 1973), providing a potential predictor of how much semantic information was carried by drawings cued by those images.

2.4.1 Participants

88 participants (42 male, mean age = 29.2 years) were recruited via Prolific. Each participant received $3.00 for their participation in the 15-minute study (approx. $12/hr). Data from 8 participants who did not meet the exclusion criteria we defined after data collection (but prior to formal analysis) were excluded from further analyses.[2] In follow-up analyses, we found that small adjustments to these exclusion criteria did not have a major impact on our key analyses.

2.4.2 Stimuli and task procedure

Each participant was presented with the prompt (*"How well does this picture fit your idea or image of the category?"*), a series of 128 images, and was asked to provide typicality judgments on a 5-point Likert scale: "Not at all", "Somewhat", "Moderately", "Very", and "Extremely." In each session, there were 4 images from each of the 32 categories. This study yielded 10,240 ratings, with each of the 1,024 photos having been rated 10 times each.

2.5 Object-part label elicitation

As in prior work that segmented freehand drawings into nameable parts (Mukherjee et al., 2019; Huey, 2022), we first elicited the names of relevant parts for each object category from an independent group of participants who were not shown any of the drawings collected in the current study.

[2]Data from an entire session were excluded if: (1) four or more catch trials out of eight were failed, (2) there were two or more response "streaks" wherein the same rating was given eight times in succession, and (3) the pattern of ratings across trials had an unusually low correlation with ratings provided by other participants.

2.5.1 Participants

150 participants (71 female, 26.15 years) were recruited from Prolific and completed the study. We did not exclude data from any participants in this sample.

2.5.2 Stimuli & Task Procedure

We collected object part labels for each of the 32 object categories, adapting a method used in prior work (Huey, 2022). On each trial, participants were cued with an object category label and asked to list 3 to 10 object parts that came to mind (e.g, tail, eye, head for the object concept of “cat”). Half of these participants were asked to list text labels that best fit their general idea of an object concept such as “cat”. The remaining participants were asked to list text labels that best fit a specific exemplar such as a specific “cat” that came to mind. Participants were instructed to write visually concrete parts of an object (e.g., nouns like “tail”) rather than abstract attributes (e.g., adjectives like “fluffy”), to use commonly known names rather than technical jargon (e.g., “stifle”), and to make a complete list of parts for each object category.

2.5.3 Data preprocessing

We applied lemmatization to the resulting part lists to remove syntactically redundant labels (e.g., “paw” vs. “paws”). We also manually edited part labels that were spelled incorrectly or semantically redundant for the object category (e.g., “fur” vs. “hair”). We then selected the top 10% most frequently listed part labels across both conditions, which provided us with a total of 340 part labels, ranging between 7-17 unique parts per object category.

2.6 Sketch part annotation

Using the part labels obtained in Section 2.5 (*Object-part label elicitation study*), we then prompted a new group of participants to tag each stroke in a series of drawings with the object part it seemed to represent, adapting methods used in prior work (Huey et al., 2023; Mukherjee et al., 2019).

Participants 6,486 participants (3,913 female, 26.4 years) were recruited from both Prolific (N=2,105) and our university study pool (N=4,381) and completed the study. We excluded data from 553 additional participants who reported experiencing technical difficulties with the web interface (N=46) and for having low accuracy on the attention-check trial, described below (N=507). Data collection stopped when every drawing had received annotations from at least three annotators.

Stimuli & Task Procedure On each trial, annotators were presented with a drawing, name of the corresponding object category (e.g., "cat"), and a menu containing the set of relevant part labels sourced in the object-part label elicitation study. Annotators were prompted to tag each stroke in a drawing with the part label that best matched it. If a stroke appeared to represent multiple different parts, they could tag it with multiple part labels. Additionally, if annotators believed that none of the provided part labels could adequately describe a stroke, they could provide their own custom label. Strokes could also be labeled as "unintelligible" if its meaning was unclear (e.g., scribbles, random dots) and/or it did not seem to depict a part of the object (e.g., grass in a drawing of a lion).

Each annotator was presented with a randomly sampled sequence of 8 drawings and one "attention-check" trial that was inserted at a randomly chosen position in the trial sequence. This "attention-check" trial contained a drawing that was considered extremely straightforward to annotate and was also annotated completely by a member of the research team. We scored these attention-check trials strictly: if an annotator failed to provide labels that exactly matched those provided by the researcher on that trial, data from that annotation session was excluded from subsequent analysis.

Data preprocessing We first manually inspected the dataset and excluded 124 drawings that were either uninterpretable (e.g., scribbles) or did not seem to be related to the target category (e.g., person drawn in the "blimp" category, lightning bolts drawn in the "ray" category, or a character from a horror story drawn in the "saw" category), yielding 12,164 remaining drawings. To evaluate how often annotators agreed on what each stroke represented in each drawing, we computed an estimate of inter-rater consistency. We found that all three annotators agreed on the same part label for 60.23% of all drawn strokes from the entire drawing dataset. We retained strokes that were assigned the same part label by at least two of three annotators. Of the 340 available part labels, we discovered that annotators only used 153 of them, suggesting that our procedure for eliciting object-part labels yielded a sufficiently broad set of options. While participants were permitted to provide custom labels, this option was very rarely used, so we did not include these custom labels in subsequent analyses of the part structure of these drawings. Strokes that were labeled as "unintelligible" were also omitted from subsequent analysis. Applying this full set of preprocessing steps yielded a total of 140,227 annotated strokes across 12,142 drawings.

3 Results

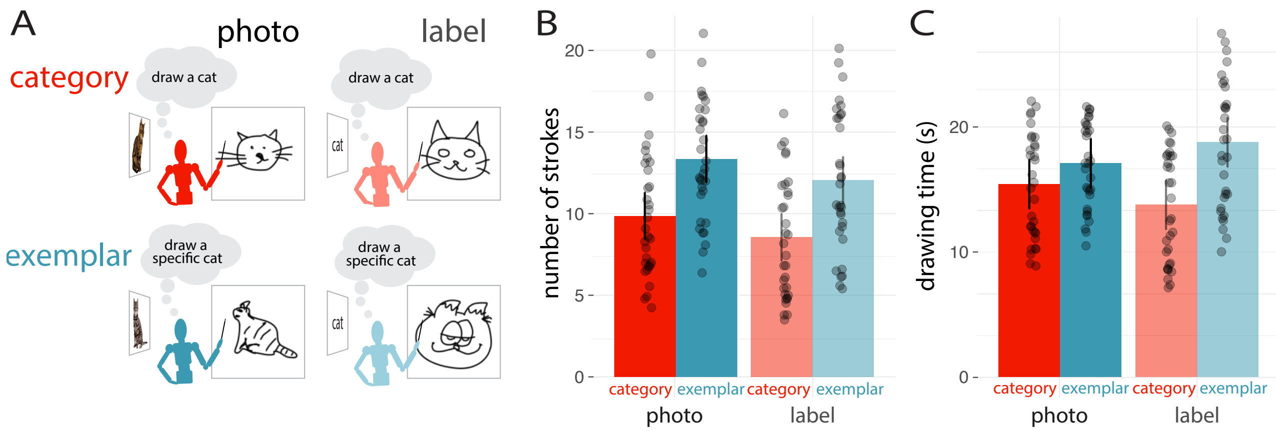

Figure 2: (A) Each participant in the sketch production study produced 32 drawings of different objects, and was assigned to one of two communicative-goal (exemplar vs. category) and cue-type conditions (photo vs. label). (B) Mean number of strokes used to produce each drawing. (C) Mean amount of time spent producing each drawing. Error bars reflect 95% CIs.

3.1 Drawings of specific objects contain more strokes and take longer to produce

Based on prior work (Fan et al., 2020), we expected drawings emphasizing distinctions between exemplars to be more detailed than those that only needed to be recognizable at the category level in order to produce sufficiently informative drawings at each level of abstraction. We further hypothesized that, insofar as the photo cues served to activate more detailed memories of what objects looked like, that photo-cued drawings would also be more detailed than label-cued ones.

To evaluate these two possibilities, we analyzed the number of strokes and amount of time participants spent producing their drawings. We first fit a linear mixed-effects model to predict the number of strokes from goal and cue type, with random intercepts for each participant and category, and found that this model outperformed nested variants of this model containing only goal ($\chi^2 = 9.8$, $p = 1.7\mathrm{e}{-3}$) or cue type ($\chi^2 = 79.0$, $p < 1\mathrm{e}{-16}$) alone. However, adding an interaction term between cue type and goal did not improve model fit ($\chi^2 = 0.11$, $p = 0.73$), suggesting that each factor independently impacted the number of strokes participants included in their drawings. Specifically, participants who had the goal of drawing an exemplar used more strokes relative to those aiming to convey a category (exemplar: 12.7 strokes, 95% CI: [11.3, 14.1]; category: 9.20 strokes, 95% CI: [7.79, 10.6]; $b = 3.49$, $t_{376} = 9.35$, $p < 2\mathrm{e}{-16}$; Fig. 2B). Moreover, participants who were cued with a photo used more strokes than those cued with a category label (photo: 11.6 strokes, 95% CI: [10.2, 13.0]; label: 10.3 strokes, 95% CI: [8.86, 11.7]; $b = 1.30$, $t_{376} = 3.49$, $p = 5.47\mathrm{e}{-3}$).

Next, we applied the same procedure to model the amount of time participants spent producing each drawing. Here we found that a mixed-effects model containing goal, cue type, and their interaction as predictors outperformed a nested variant lacking the interaction term ($\chi^2 = 6.22$, $p = 0.0126$). Further inspection of the coefficients of each term revealed that having the exemplar goal (exemplar: 17.9 seconds, 95% CI: [16.2, 19.7]; category: 14.6 seconds, 95% CI: [12.8, 16.4]; $b = 1.66$, $t_{378} = 1.76$, $p = 0.0788$) and being cued with a photo led to modest increases in the amount of time participants spent producing their drawing (photo: 16.3 seconds, 95% CI: [14.5, 18.0]; label: 16.3 seconds, 95% CI: [14.5, 18.0]; $b = 1.65$, $t_{378} = 1.75$, $p = 0.0808$), with the effect of goal being larger in the label-cue context (exemplar: 18.8 seconds, 95% CI: [16.8, 20.8]; category: 13.8 seconds, 95% CI: [11.8, 15.8]; $b = 3.33$, $t_{378} = 2.49$, $p = 0.0130$; Fig. 2C).

These findings suggest that both participants' goal and the availability of a reference image impacted the level of detail in participants' drawings, replicating and extending prior work that had used a more indirect manipulation of task goals but did not manipulate visual access (Fan et al., 2020).

3.2 Drawings of specific objects are less categorizable than drawings intended to communicate category information

We next sought to measure the impact of having different goals during drawing production on the amount of category evidence in a drawing, as measured by how well naive participants could determine the category a drawing was intended to convey. We used a mixed-effects logistic regression model to predict the outcome of each categorization judgment using goal, cue type, as well as their interaction, with random intercepts for drawing participant, categorization participant, and target category. This model outperformed nested variants that did not include the interaction term ($\chi^2 = 7.13$, $p = 0.008$) or excluded either goal ($\chi^2 = 11.7$, $p = 6.3\text{e}{-4}$) or cue type ($\chi^2 = 28.7$, $p = 8.5\text{e}{-8}$), suggesting that both factors influenced the categorizability of a drawing. To control for potential differences in recognizability due to individual differences in drawing skill and systematic differences in the amount of detail in drawings across conditions, we next considered an augmented model containing two additional covariates: self-reported drawing skill and the number of strokes in each drawing. We found that this augmented model outperformed the original one without these covariates by a large margin ($\chi^2 = 62.3$, $p = 2.9\text{e}{-14}$), suggesting that these additional variables account for meaningful additional variance in drawing categorizability.

Further examination of the model coefficients revealed that exemplar drawings were less categorizable than category drawings (exemplar: 0.802, 95% CI: [0.739, 0.854], category: 0.847, 95% CI: [0.794, 0.888]; $b = -0.48$, $z = -4.88$, $p = 1.1\text{e}{-6}$), photo-cued drawings were less categorizable than label-cued drawings

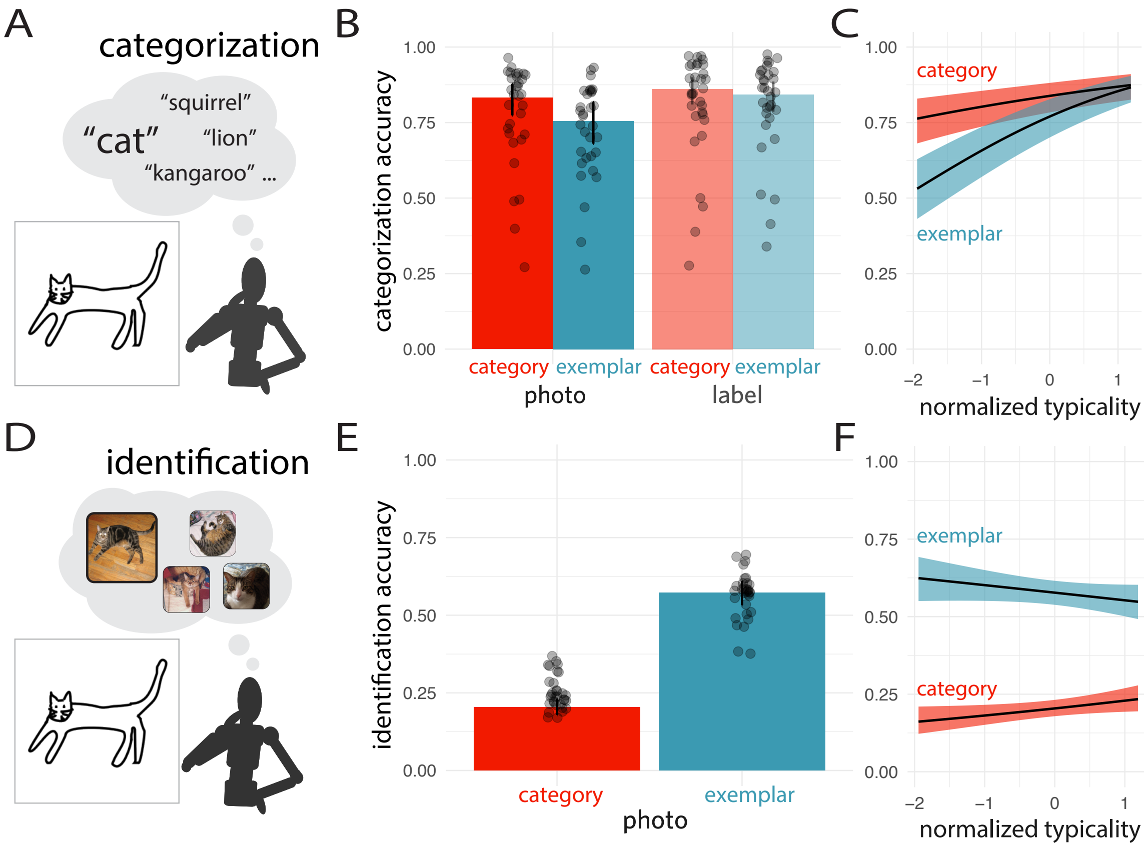

Figure 3: (A) Participants in the sketch categorization study selected the category label that best matched the drawing from the full set of 32 category labels. (B) Proportion of drawings correctly categorized across conditions. Error bars reflect 95% CIs. (C) Categorization accuracy for photo-cued drawings as a function of the typicality of the photo cue. Error ribbons reflect 95% CIs. (D) Participants in the sketch identification study selected the photo that best matched the drawing from a set of 8 photos belonging to the target category. Only photo-cued drawings were included in this experiment. (E) Proportion of photo-cued drawings correctly matched to their corresponding photo cue. (F) Exemplar identification accuracy for photo-cued drawings as a function of the typicality of the photo cue.

(photo: 0.797, 95% CI: [0.732, 0.849]; label: 0.852, 95% CI: [0.800, 0.892]; $b = 0.22$, $z = -2.29$, $p = 0.026$), and the gap between exemplar and category drawings was more pronounced in the photo-cue condition ($b = 0.33$, $z = 2.36$, $p = 0.018$; Fig. 3B). These results show that goal and cue type interact to impact the amount of category evidence in a drawing. Specifically, drawings intended to depict a specific exemplar that was just seen contain less category-diagnostic information than drawings intended to evoke the general category. However, when participants are prompted with only a category label, the effect of goal is more modest.

The finding that photo-cued drawings were less recognizable at the category level merited further investigation given a previous study that had varied cue type found the opposite pattern of results (Fan et al., 2018a), with photo-cued drawings outperforming label-cued ones. That finding had been interpreted as potentially reflecting the ability of photos to remind participants of category-diagnostic visual details that may have otherwise been

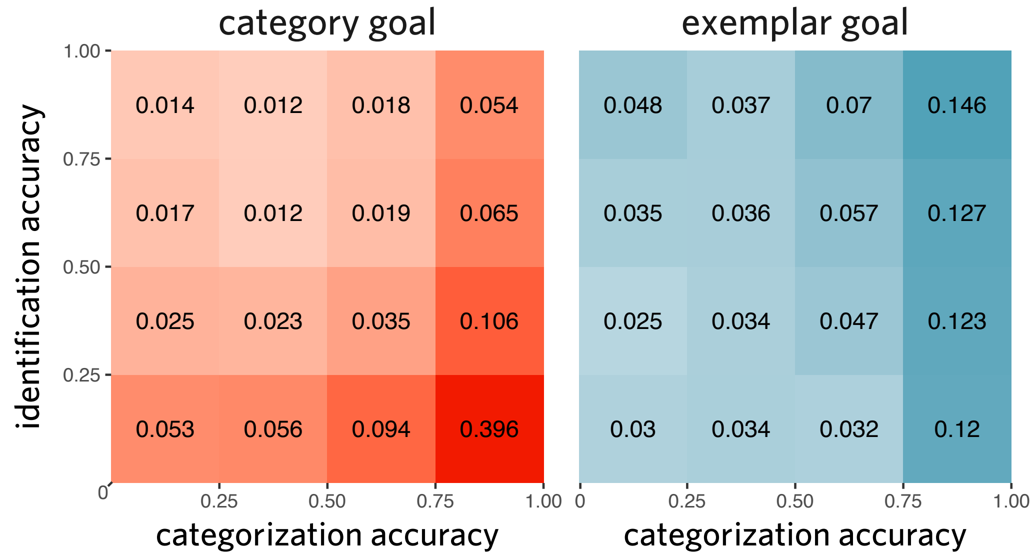

Figure 4: Joint frequency distribution of photo-cued drawings achieving different levels of categorization and exemplar identification accuracy under the category goal (left) and exemplar goal (right).

difficult to retrieve from long-term knowledge. Towards reconciling these two sets of findings, we explored the possibility that the degree to which photo cues impact a drawing's categorizability depends on how prototypical the photographed exemplar is. Specifically, less typical photos might yield less categorizable drawings because these photos contain less category-diagnostic information for participants to draw upon. To evaluate this possibility, we fit categorization judgments for photo-cued drawings with a mixed-effects logistic regression model containing both cue typicality, goal, and their interaction as fixed effects, and random intercepts for drawing participant, categorization participant, and target category. We found that this model outperformed reduced variants lacking the interaction term ($\chi^2 = 18.5$, $p = 1.7e{-5}$), goal ($\chi^2 = 17.6$, $p = 2.7e{-5}$), and typicality ($\chi^2 = 129.4$, $p < 2e{-16}$) as predictors, suggesting that all three terms explained meaningful amounts of variation in categorization performance. Inspection of the coefficients of this model revealed that, indeed, less typical photo cues yielded less categorizable drawings ($b = -0.25$, $z = 4.62$, $p = 3.9e{-6}$), with the effect of typicality being greater for exemplar drawings ($b = 0.31$, $z = 4.34$, $p = 1.44e{-4}$; Fig. 3C). Together, these findings suggest a more nuanced view of when photo cues help people produce drawings that are easier to categorize, and when they do not.

3.3 Drawings of specific objects are more identifiable, even at the expense of categorizability

In the previous section, we found that participants intending to depict an exemplar, especially a less typical one, produce drawings that are less recognizable at the category level. One possible explanation for these results is that the task of faithfully capturing the visual appearance of these exemplars was sufficiently challenging that participants simply failed to encode meaningful semantic information in their drawings at all, whether at the exemplar or category levels. Under this account, exemplar drawings would also underperform on the exemplar identification task. Alternatively, these drawings may have succeeded in encoding the distinctive visual details that support recognition at the exemplar level, leading them to outperform category drawings on this recognition task.

To tease apart these possibilities, we constructed mixed-effects logistic regression models with the same random effects structure as above to analyze how easily photo-cued drawings could be matched with their corresponding photo cue. We found using nested model comparison that goal had a large impact on exemplar identification ($\chi^2 = 158.4$, $p < 2\text{e}{-16}$), with exemplar drawings achieving reasonably high accuracy and substantially outperforming category drawings (exemplar: 0.572, 95% CI: [0.532, 0.611]; category: 0.204, 95% CI: [0.179, 0.231]; $b = 1.65$, $z = 15.4$, $p < 2\text{e}{-16}$; Fig. 3E). When we augmented this model with information about cue typicality, we discovered that typicality had an impact on exemplar identification, but in different ways depending on goal ($b = 0.248$, $z = 3.53$, $p = 4.2\text{e}{-4}$; Fig. 3F). Specifically, while category drawings were more easily matched to their corresponding photo cue when the cue was more typical, exemplar drawings were more difficult to match to more typical photo cues. This pattern of results may reflect the shifting demands of the exemplar-recognition task depending on the typicality of the photo cue: for category drawings, this task is more like a category recognition task when the cue is more typical; for exemplar drawings, this task may be more challenging for more typical photo cues because there are fewer distinguishing features that the drawing and photo can share that are not also shared by the distractor images.

To further explore a potential dissociation between category and exemplar information in drawings, we next modeled the relationship between the categorization accuracy and identification accuracy achieved by an individual photo-cued drawing (Fig. 4). We observed that category drawings that were the most categorizable (i.e., above 75% correct) were also the most difficult to match to their corresponding photo cue (i.e., below 25% correct). On the other hand, while many exemplar drawings still achieved relatively high categorization accuracy, among those that were the most difficult to categorize (i.e., below 25% correct), a meaningful proportion achieved higher identification accuracy. Motivated by these observations, we formally evaluated evidence for a trade-off by constructing a linear mixed-effects regression model predicting identification accuracy for individual drawings from categorization accuracy, goal, and their interaction, with random intercepts for drawing-production participant and target category. This analysis revealed evidence for a reliable negative relationship between categorization accuracy and identification accuracy ($b = -0.0695$, $t = -3.40$, $p = 6.9\text{e}{-4}$) that did not depend on goal condition. Taken together, these additional findings provide additional support for the notion that category-level and exemplar-specific information can trade off to produce different visual recognition patterns.

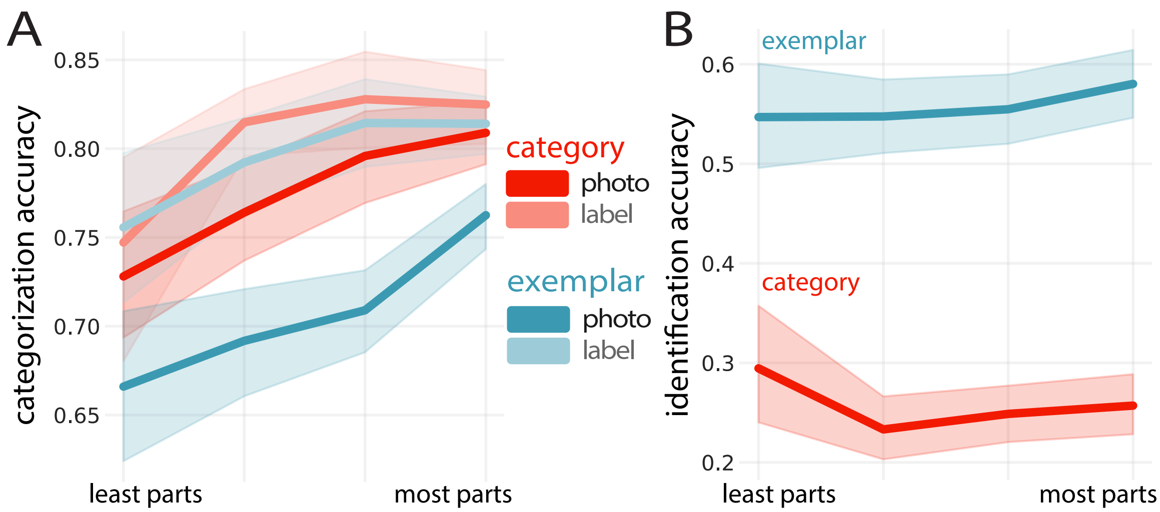

Figure 5: (A) Categorization accuracy as a function of the number of unique object parts, divided into quartiles. (B) Identification accuracy as a function of the number of unique object parts, divided into quartiles.

3.4 Drawings of specific objects contain more parts, especially when cued by a category label

The findings so far indicate that drawings produced under different task conditions drive different visual recognition patterns, raising the question as to what features of these drawings might underlie those effects. To address this question, we leveraged stroke-level part annotations to analyze the semantic part complexity of these drawings. Specifically, we asked to what degree drawings produced in some conditions contained more nameable parts than in the others (regardless of how many strokes were used to depict each part). To answer this question, we constructed a linear mixed-effects model predicting the number of nameable parts in a drawing from goal, cue type, and their interaction, with random intercepts for target category. We found that this model outperformed nested variants that omitted the interaction term ($\chi^2 = 19.37$, $p = 1.08\text{e}{-5}$), excluded goal ($\chi^2 = 119.79$, $p < 2.2\text{e}{-16}$), or excluded cue type ($\chi^2 = 7.72$, $p = 0.0055$), suggesting that both factors impacted the part complexity of the resulting drawings. In particular, we found that exemplar drawings reliably contained a slightly greater number of nameable parts than category drawings (exemplar: 6.68; category: 6.30; 95% CI: $[6.02, 7.34]$ and $[5.63, 6.96]$, respectively), suggesting that drawings intended to convey a more specific meaning tend to display greater semantic part complexity. Similarly, photo-cued drawings contained more parts than label-cued ones (photo: 6.54; text: 6.44; 95% CI: $[5.87, 7.20]$ and $[5.78, 7.10]$, respectively), suggesting that producing a drawing just after having seen an exemplar of it can also increase its part complexity. However, the interaction between goal and cue type was such that exemplar drawings prompted with a label cue contained more parts than those prompted with a photo cue (label-exemplar: 6.71; photo-exemplar: 6.65; 95% CI:

$[6.04, 7.37]$ and $[5.99, 7.32]$, respectively), suggesting that participants who were cued only the category label recalled a greater diversity of nameable attributes when trying to draw an exemplar, by comparison with those who were provided with an image cue. Taken together, these findings provide converging evidence that our manipulation of goal and cue type in the sketch production study reliably affected the semantic part complexity of the resulting drawings.

3.5 Drawings that contain more parts are more categorizable, but not necessarily more identifiable

Next, we sought to evaluate the potential link between a drawing's semantic part complexity and how recognizable it is at both the category and exemplar levels. Towards this end, we first constructed a mixed-effects logistic regression model predicting the odds of successful categorization from the number of nameable parts, goal, cue type, and the interaction between goal and cue type, with random intercepts for target category. We found that this model outperformed a nested variant that omitted the number of nameable parts ($\chi^2 = 51.34$, $p = 7.79\text{e}{-13}$), suggesting that the semantic part complexity of a drawing is a reliable predictor of its categorizability, even after controlling for experimental condition. Specifically, we found that each additional nameable part that a drawing included was associated with increased categorization accuracy ($\beta = 0.0077$, $t = 7.17$, $p = 7.88 \times 10^{-13}$). For instance, predicted categorization accuracy increased from 0.74 (95% CI: [0.72, 0.75]) when a drawing contained only one nameable part to 0.86 (95% CI: [0.84, 0.88]) when it contained 17 parts. Next, for photo-cued drawings only, we constructed an analogous mixed-effects logistic regression model predicting the odds of successful *identification*. This analysis revealed that drawings containing more parts were not more often matched to its corresponding photo cue ($\chi^2 = 0.72$, $p = 0.395$), suggesting that variation in identifiability across drawings is not well explained by how many distinct parts a drawing contains. Taken together, these exploratory analyses suggest that the semantic part complexity of a drawing is predictive of its categorizability, but not necessarily how easily it can be identified.

3.6 Drawings of specific objects and drawings of object categories emphasize different parts

While the previous section analyzed the semantic content of drawings only in terms of *how many* parts they contained, having detailed part annotations for each drawing enabled further exploration of *which* parts were contained in drawings in different conditions, and *to what degree they were emphasized*. Towards this end, we represented each drawing in two ways, following prior work (Mukherjee et al., 2019). The first was a binary vector representation, where each element represented whether a specific part was present or absent, such that

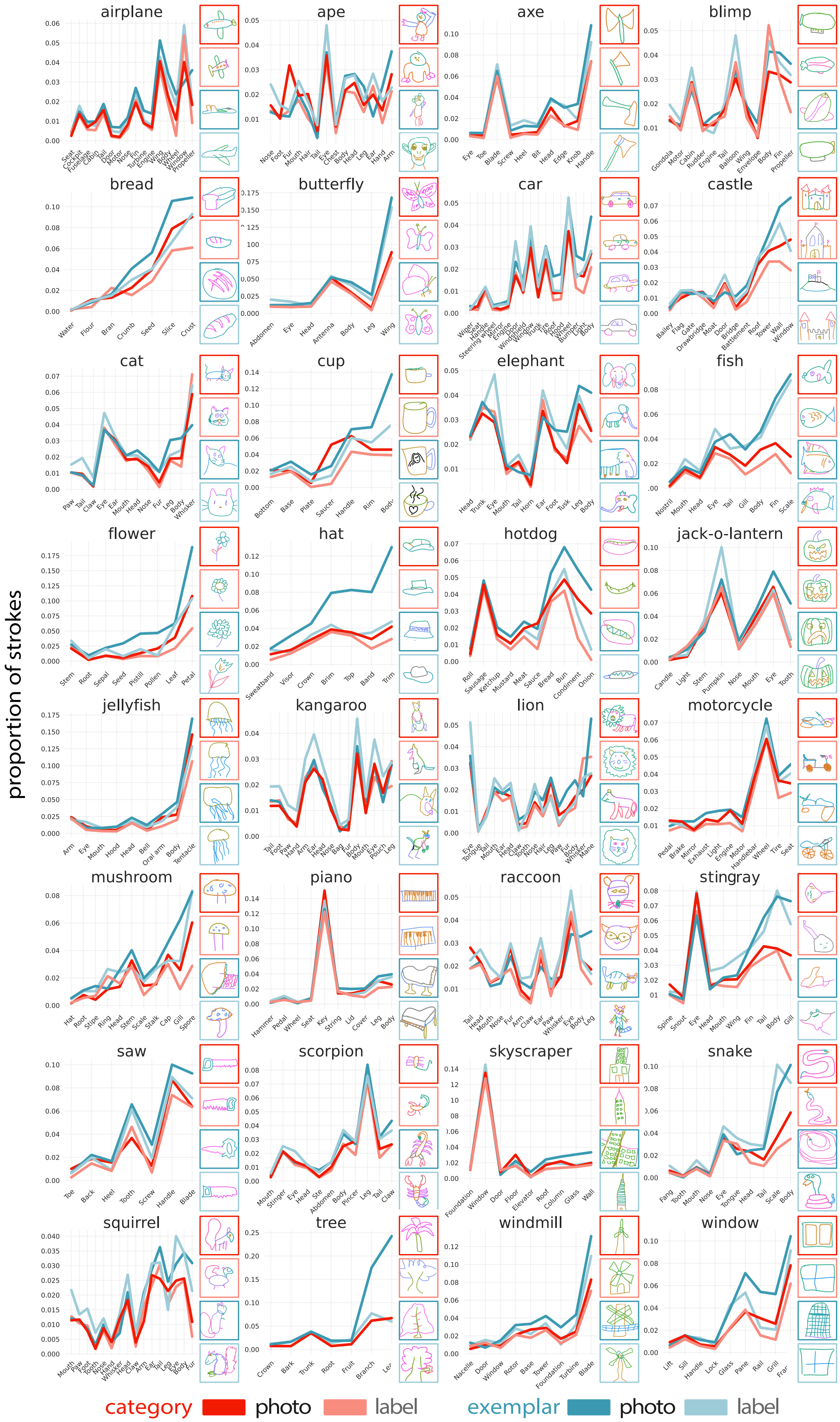

Figure 6: The proportion of strokes allocated to different object parts in each sketch production condition. Example drawings are shown with each part rendered in a unique color.

one of these vectors encodes the full set of parts that an individual drawing contained. The second is a continuous vector representation, where each element represented the proportion of the number of strokes in a drawing used to depict a specific part, such that one of these vectors can be thought of as encoding the pattern of emphasis on different parts of an object (Figure 6). We then used a permutation test to assess how reliably drawings produced in each condition differed from each other in terms of these two ways of representing their semantic part content. Separately for each category, we shuffled the assignment of drawings to condition 5,000 times and computed the average difference between the vector representations for each condition each time, yielding a null distribution for the difference between conditions. We used a metric known as Jensen–Shannon divergence (JSD)–commonly used for quantifying the divergence between two probability distributions–to measure the difference between these averaged vector representations. We then derived a p-value by computing the proportion of all shuffled iterations that yielded JSD values greater than or equal to the observed JSD value for each category. This analysis suggested that participants emphasized parts differently when producing a photo-cued exemplar drawing than when producing a label-cued category drawing ($p = 0.030$), and provided marginal evidence that photo-cued exemplar drawings and label-cued category drawings contained reliably distinct sets of parts ($p = 0.051$). There were no other reliable differences between conditions. Taken together, this exploratory non-parametric analysis suggests that different task conditions can lead to drawings that emphasize object parts in different ways, with potential implications for downstream recognition behavior.

4 Discussion

Here we investigated how people visually communicate about object concepts across different task conditions. Using a drawing paradigm, we independently manipulated whether the goal was to produce a drawing of a specific exemplar or a general category, and whether participants had visual access to a photograph of an object just before starting to draw. Our core question concerned differences in the semantic information carried by drawings produced under these settings. Specifically, are drawings that look more like a specific exemplar necessarily easier to categorize, as well? Or can information at the category and exemplar level trade off against one another, such that when people produce drawings that more strongly evoke a category, these drawings are less evocative of specific exemplars, and vice versa? Our experiments provide evidence in favor of such a trade-off, such that exemplar drawings are easier to identify but less categorizable than category drawings. Moreover, drawings that were cued by photograph are less categorizable than drawings cued by category label, although this gap is reduced when the photograph is of a more typical exemplar. Further exploration of the internal semantic structure of these drawings suggests that these differences in recognizability might be mediated, at least in part, by how many distinct parts different drawings contain and variation in the pattern of emphasis on different parts. Taken together, our data provide a more nuanced understanding of how drawings encode meaning at different levels of semantic abstraction, suggesting a dissociation between how drawings communicate more general and more specific meanings. In addition, another contribution we aim to make in this work is to provide a publicly available dataset containing freehand drawings spanning a diversity of object concepts, as well as detailed annotations of the part structure of these drawings.

Although we interpret differences between the photo-cue and label-cue conditions as being primarily driven by differences in modality (i.e., visual vs. linguistic), the two cue conditions also varied in other ways. Specifically, we compared the impact of using a photo cue (accompanied by a category label) to that of using only a basic-level category label, and did not include a condition where the text-based cue referred to a subordinate category or described an exemplar (Rosch, 1973; Bauer and Just, 2017). As such, the photo-cue and label-cue conditions also differed in the level of semantic abstraction at which the visual object concept was activated. Thus it is not clear to what degree the differences we measured across conditions reflect differences in modality, *per se*, as opposed to differences in the level of semantic abstraction. Future work could bring greater clarity to this issue by including a condition in which drawings are cued by more specific and detailed descriptions compared with drawings cued by photos.

The present study focused on real-world objects that we expected participants in our samples to generally be familiar with, allowing them to rely on pre-existing semantic knowledge when producing drawings or recognizing them. As such, the degree to which the trade-off between category-level and exemplar-specific information is dependent on such prior knowledge is not clear. For example, it might be that drawings of a "generic" airplane systematically differ from drawings of any specific airplane by emphasizing features that people have gradually learned to treat as diagnostic. The role of culturally transmitted conventions for depicting visual concepts might be one source of such consistency, along with direct visual experience with these objects. One promising direction for more directly characterizing the role of direct visual experience would be to conduct a replication of the current study with novel objects. In such a study, participants would be unable to rely on pre-existing semantic knowledge, yet could still selectively communicate about these objects at different levels of abstraction.

Here we employed crowdsourcing to characterize the visual recognition patterns evoked by drawings, as well as to obtain detailed annotations of the part structure of each drawing. A natural follow-up question, given recent advances in computational vision (Yamins et al., 2014; Hong et al., 2016; Zhuang et al., 2021), is to develop more detailed accounts of the visual computations supporting human sketch understanding. Prior work has used artificial neural networks to model some aspects of how humans extract category-level information from drawings (Fan et al., 2018a). However, the models used in that work may be less well suited to representing fine-grained visual differences between exemplars. Thus another natural direction future work would be to evaluate a broader array of computational models that embody different hypotheses about the underlying representation and prior experience required to support recognition of drawings at both the category and exemplar levels (Fan et al., 2020; Mukherjee et al., 2023), leveraging more recently developed models optimized for exemplar discrimination (Vinker et al., 2022; Wu et al., 2018; Zhuang et al., 2021).

In sum, our findings contribute to a growing body of work using drawings to investigate various aspects of cognition, including learning (Fan et al., 2018a; Fiorella and Zhang, 2018; Chamberlain et al., 2021), communication (Hawkins et al., 2019; Fan et al., 2020; Garrod et al., 2007), memory (Bainbridge et al., 2019; Roberts and Wammes, 2020; Wammes et al., 2016), and development (Dillon, 2020; Long et al., 2024; Kellogg, 1969). Such approaches highlight the value of using open-ended production tasks to gain insight into what people perceive and know about the visual world (Fan et al., 2023).

5 Constraints on Generality

Participants in all studies, except those in Section 2.1, were adult English speakers recruited online using Prolific. The study described in Section 2.1 also recruited participants from a U.S. university participant pool, primarily young adult English speakers (ages 18–22). These populations served as the focus for the current work because they facilitate comparison with prior experimental work that has investigated drawing production and recognition in comparable controlled settings. As such, additional studies are needed to assess the generalizability of our findings to participant populations that are not as easily accessible via online crowdsourcing platforms. In addition, because our study elicited drawings using a digital drawing canvas, where many participants used either a mouse or trackpad to produce their drawings, further work is also needed to assess the generalizablity of our findings to contexts where people draw using a stylus and where people are drawing on paper, rather than a digital canvas.

6 Acknowledgments

Masked for review.

7 Author Contributions

Masked for review.

References

Aubert, M., Brumm, A., Ramli, M., Sutikna, T., Saptomo, E. W., Hakim, B., Morwood, M. J., van den Bergh, G. D., Kinsley, L., and Dosseto, A. (2014). Pleistocene cave art from sulawesi, indonesia. *Nature*, 514(7521):223–227.

Bainbridge, W. A., Hall, E. H., and Baker, C. I. (2019). Drawings of real-world scenes during free recall reveal detailed object and spatial information in memory. *Nature Communications*, 10(1):1–13.

Bauer, A. J. and Just, M. A. (2017). A brain-based account of "basic-level" concepts. *Neuroimage*, 161:196–205.

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

Boutonnet, B. and Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*, 35(25):9329–9335.

Bozeat, S., Lambon Ralph, M. A., Graham, K. S., Patterson, K., Wilkin, H., Rowland, J., Rogers, T. T., and Hodges, J. R. (2003). A duck with four legs: Investigating the structure of conceptual knowledge using picture drawing in semantic dementia. *Cognitive neuropsychology*, 20(1):27–47.

Chamberlain, R., Kozbelt, A., Drake, J. E., and Wagemans, J. (2021). Learning to see by learning to draw: A longitudinal analysis of the relationship between representational drawing training and visuospatial skill. *Psychology of Aesthetics, Creativity, and the Arts*, 15(1):76.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., and Goodman, N. D. (2020). When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychological Review*, 127(4):591.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Dillon, M. R. (2020). Rooms without walls: Young children draw objects but not layouts. *Journal of Experimental Psychology: General*.

Edelman, S. et al. (1999). *Representation and recognition in vision*. MIT press.

Eitz, M., Hays, J., and Alexa, M. (2012). How do humans sketch objects? *ACM Transactions on graphics (TOG)*, 31(4):1–10.

Fan, J., Yamins, D., and Turk-Browne, N. (2018a). Common object representations for visual production and recognition. *Cognitive Science*, 42:2670–2698.

Fan, J. E., Bainbridge, W. A., Chamberlain, R., and Wammes, J. D. (2023). Drawing as a versatile cognitive tool. *Nature Reviews Psychology*, 2(9):556–568.

Fan, J. E., Hawkins, R. D., Wu, M., and Goodman, N. D. (2020). Pragmatic inference and visual abstraction enable contextual flexibility during visual communication. *Computational Brain & Behavior*, 3(1):86–101.

Fan, J. E., Yamins, D. L., and Turk-Browne, N. B. (2018b). Common object representations for visual production and recognition. *Cognitive science*, 42(8):2670–2698.

Fiorella, L. and Zhang, Q. (2018). Drawing boundary conditions for learning by drawing. *Educational Psychology Review*, 30(3):1115–1137.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., and Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive neuropsychology*, 18(2):125–174.

Garrod, S., Fay, N., Lee, J., Oberlander, J., and MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, 31(6):961–987.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In *Perceptual organization in vision*, pages 245–290. Psychology Press.

Hawkins, R. X., Sano, M., Goodman, N. D., and Fan, J. W. (2019). Disentangling contributions of visual information and interaction history in the formation of graphical conventions. In *CogSci*, pages 415–421.

Hoffmann, D. L., Standish, C. D., García-Diez, M., Pettitt, P. B., Milton, J. A., Zilhão, J., Alcolea-González, J. J., Cantalejo-Duarte, P., Collado, H., de Balbín, R., Lorblanchet, M., Ramos-Muñoz, J., Weniger, G.-C., and Pike, A. W. G. (2018). U-th dating of carbonate crusts reveals neandertal origin of iberian cave art. *Science*, 359(6378):912–915.

Hong, H., Yamins, D. L., Majaj, N. J., and DiCarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature neuroscience*, 19(4):613–622.

Huey, H. (2022). Developmental changes in the semantic part structure of drawn objects. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*.

Huey, H., Lu, X., Walker, C. M., and Fan, J. E. (2023). Visual explanations prioritize functional properties at the expense of visual fidelity. *Cognition*, 236:105414.

Kellogg, R. (1969). *Analyzing children's art*. National Press Books Palo Alto, CA.

Logothetis, N. K. and Sheinberg, D. L. (1996). Visual object recognition. *Annual review of neuroscience*, 19(1):577–621.

Long, B., Fan, J. E., Huey, H., Chai, Z., and Frank, M. C. (2024). Parallel developmental changes in children's production and recognition of line drawings of visual concepts. *Nature Communications*, 15(1):1191.

Lupyan, G. and Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, 141(1):170.

McClelland, J. L. and Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature reviews neuroscience*, 4(4):310–322.

McClelland, J. L. and Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114(2):159.

McRae, K., Cree, G. S., Seidenberg, M. S., and McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, 37(4):547–559.

Medin, D. L. and Schaffer, M. M. (1978). Context theory of classification learning. *Psychological review*, 85(3):207.

Mukherjee, K., Hawkins, R. X., and Fan, J. W. (2019). Communicating semantic part information in drawings. In *CogSci*, pages 2413–2419.

Mukherjee, K., Lu, X., Huey, H., Vinker, Y., Aguina-Kang, R., Shamir, A., and Fan, J. E. (2023). Evaluating machine comprehension of sketch meaning at different levels of abstraction. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 45.

Murphy, G. (2004). *The big book of concepts*. MIT press.

Murphy, G. L. and Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of verbal learning and verbal behavior*, 21(1):1–20.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, 115(1):39.

Palmeri, T. J. and Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5(4):291–303.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory*, 2(5):509.

Roberts, B. R. and Wammes, J. D. (2020). Drawing and memory: Using visual production to alleviate concreteness effects. *Psychonomic Bulletin & Review*, pages 1–9.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., and Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological review*, 111(1):205.

Rosch, E. and Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4):573–605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3):382–439.

Rosch, E. H. (1973). Natural categories. *Cognitive psychology*, 4(3):328–350.

Sangkloy, P., Burnell, N., Ham, C., and Hays, J. (2016). The sketchy database: Learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (proceedings of SIGGRAPH)*.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Suresh, S., Mukherjee, K., Giallanza, T., Yu, X., Patil, M., Cohen, J. D., and Rogers, T. T. (2025). Ai-enhanced semantic feature norms for 786 concepts. *arXiv preprint arXiv:2505.10718*.

Tanaka, J. W. and Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive psychology*, 23(3):457–482.

Tversky, B. and Hemenway, K. (1984). Objects, parts, and categories. *Journal of experimental psychology: General*, 113(2):169.

Tylén, K., Fusaroli, R., Rojo, S., Heimann, K., Fay, N., Johannsen, N. N., Riede, F., and Lombard, M. (2020). The evolution of early symbolic behavior in homo sapiens. *Proceedings of the National Academy of Sciences*, 117(9):4578–4584.

Vinker, Y., Pajouheshgar, E., Bo, J. Y., Bachmann, R. C., Bermano, A. H., Cohen-Or, D., Zamir, A., and Shamir, A. (2022). Clipasso: Semantically-aware object sketching. *arXiv preprint arXiv:2202.05822*.

Vinson, D. P. and Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40(1):183–190.

Wammes, J. D., Meade, M. E., and Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *Quarterly Journal of Experimental Psychology*, 69(9):1752–1776.

Wu, L.-l. and Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta psychologica*, 132(2):173–189.

Wu, Z., Xiong, Y., Yu, S. X., and Lin, D. (2018). Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and 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.

Yang, J. and Fan, J. E. (2021). Visual communication of object concepts at different levels of abstraction. *arXiv preprint arXiv:2106.02775*.

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., and Yamins, D. L. (2021). Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, 118(3):e2014196118.