Disentangling contributions of visual information and interaction history in the formation of graphical conventions

Robert X. D. Hawkins*
Department of Psychology
Stanford University
rxdh@stanford.edu

Megumi Sano*
Department of Psychology
Stanford University
megsano@stanford.edu

Noah D. Goodman
Department of Psychology
Stanford University
ngoodman@stanford.edu

Judith E. Fan
Department of Psychology
UC San Diego
jefan@ucsd.edu

Abstract

Drawing is a versatile technique for visual communication, ranging from photorealistic renderings to schematic diagrams consisting entirely of symbols. How does a medium spanning such a broad range of appearances reliably convey meaning? A natural possibility is that drawings derive meaning from both their visual properties as well as shared knowledge between people who use them to communicate. Here we evaluate this possibility in a drawing-based reference game in which two participants repeatedly communicated about visual objects. Across a series of controlled experiments, we found that pairs of participants discover increasingly sparse yet effective ways of depicting objects. These gains were specific to those objects that were repeatedly referenced, and went beyond what could be explained by task practice or the visual properties of the drawings alone. We employed modern techniques from computer vision to characterize how the high-level visual features of drawings changed, finding that drawings of the same object became more consistent within a pair of participants and divergent across participants from different interactions. Taken together, these findings suggest that visual communication promotes the emergence of depictions whose meanings are increasingly determined by shared knowledge rather than their visual properties alone.

Keywords: alignment; coordination; iconicity; sketch understanding; visual communication

Introduction

From ancient etchings on cave walls to modern digital displays, visual communication lies at the heart of key human innovations (e.g., cartography, data visualization) and forms a durable foundation for the cultural transmission of knowledge and higher-level reasoning. Perhaps the most basic and versatile technique supporting visual communication is drawing, the earliest examples of which date to at least 40,000-60,000 years ago (Hoffmann et al., 2018). What began as simple mark making has since been adapted to a wide array of applications, ranging from photorealistic rendering to schematic diagrams consisting entirely of symbols.

Even in the relatively straightforward case of drawing from life, there are countless ways to depict the same object. How does a communication medium spanning such a broad range of appearances reliably convey meaning? On the one hand, prior work has found that semantic information in a figurative drawing, i.e., the object it represents, can be derived purely from its visual properties (Fan, Yamins, & Turk-Browne, 2018). On the other hand, other work has emphasized the role of socially-mediated information for making appropriate inferences about what even a figurative drawing represents (Goodman, 1976).

How can these two perspectives be reconciled? Our approach is to consider the joint contributions of visual information and social context in determining how drawings derive meaning (Abell, 2009), and to propose that a critical factor affecting the balance between the two may be the amount of shared knowledge between communicators. Specifically, we explore the hypothesis that accumulation of shared knowledge via extended visual communication may promote the development of increasingly schematic yet effective ways of depicting an object, even as these *ad hoc* graphical conventions may be less readily apprehended by others who lack this shared knowledge.

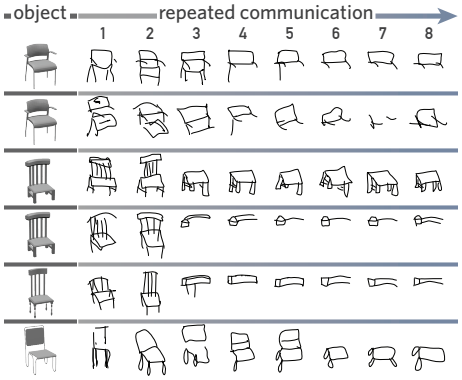

Figure 1: Repeated visual communication depicting the same object.

To investigate this hypothesis, we used an interactive drawing-based reference game in which two participants repeatedly communicated about visual objects. We examined both how their task performance and the drawings they produced changed over time (see Fig. 1). Our approach was inspired by a large literature that has explored how extended interaction influences communicative behavior in several modalities, including language (Clark & Wilkes-Gibbs, 1986; Hawkins, Frank, & Goodman, 2017), gesture (Goldin-Meadow, McNeill, & Singleton, 1996), and drawings (Garrod, Fay, Lee, Oberlander, & MacLeod, 2007; Galantucci, 2005). There are three aspects of the current work that advance our prior understanding: *first*, we include a control set of objects that were not repeatedly drawn but only shown at the beginning and end of the interaction, allowing us to measure the specific contribution of repeated reference vs. general practice effects; *second*, we measure how strongly the visual properties of drawings drive recognition in the absence of interaction history for naive viewers, while equating other task variables; and *third*, we employ recent advances in computer vision to quantitatively characterize changes in the high-level visual properties of drawings across repetitions.

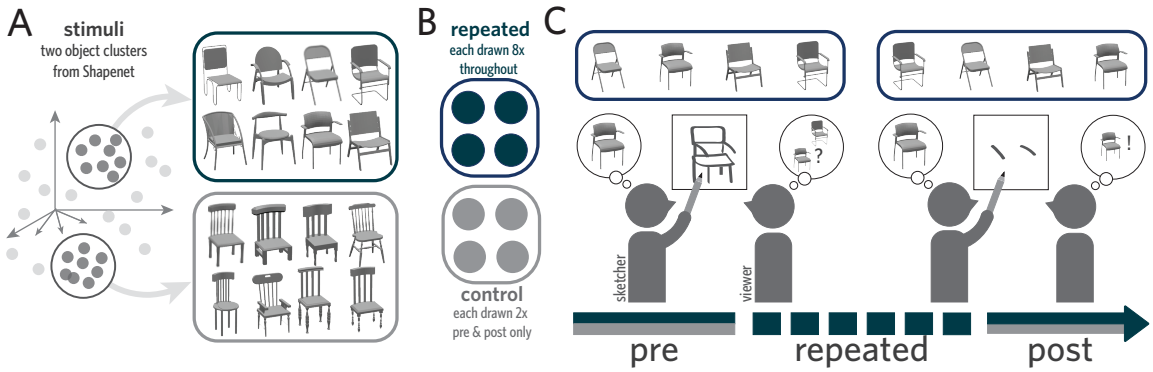

Figure 2: (A) Stimuli from ShapeNet. (B) Each pair of participants was randomly assigned two sets of four objects, each set from one of the two categories. (C) Repeated objects drawn eight times throughout; control objects drawn once at the beginning and end of each interaction.

Part I: How does repeated reference support successful visual communication?

Our first goal was to understand how people learn to communicate about visual objects across repeated visual communication. To accomplish this, we developed a drawing-based reference game for two participants. On each trial, both participants shared a *communicative context*, represented by an array of four objects. One of these objects was privately designated the 'target' to the sketcher. The sketcher's goal was to draw the target so that the viewer could select it from the array as quickly and accurately as possible. We hypothesized that learning would be *object-specific*: that over repeated visual reference to a particular object, participants would discover ways of depicting that object more effectively relative to non-repeated control objects.

Methods: Visual communication experiment

Participants We recruited 138 participants from Amazon Mechanical Turk, who were grouped into 69 pairs (Hawkins, 2015). Within each experimental session, one participant was assigned the sketcher role and the other the viewer role, and these role assignments remained the same throughout the experiment. Data from two pairs were excluded due to unusually low performance (i.e., accuracy $<$ 3 s.d. below the mean). In this and subsequent experiments, participants provided informed consent in accordance with the Stanford IRB.

Stimuli In order to make our task sufficiently challenging, we sought to construct communicative contexts consisting of objects whose members were both geometrically complex and visually similar. To accomplish this, we sampled objects from the ShapeNet (Chang et al., 2015), a database containing a large number of 3D mesh models of real-world objects. We restricted our search to 3096 objects belonging to the `chair` class, which is among the most diverse and abundant in ShapeNet. To identify groups of visually similar chairs, we first extracted high-level visual features from 2D renderings of each object using a deep convolutional neural network (DCNN) architecture, VGG-19 (Simonyan & Zisserman, 2014). This network had been previously trained to recognize objects in photos from the ImageNet database (Deng et al., 2009), containing 1.2 million natural photographs of 1000 different object classes. Trained DCNN models have been shown to predict human perceptual similarity judgments about objects (Kubilius, Bracci, & de Beeck, 2016; Peterson, Abbott, & Griffiths, 2018), as well as neural population responses in visual cortex during object recognition (Yamins et al., 2014; Güçlü & van Gerven, 2015). As such, they provide a principled choice of encoding model for extracting high-level visual information from images. Following previous work that has employed DCNN models to evaluate perceptual similarity (Peterson et al., 2018; Kubilius et al., 2016), for each image we extract a 4096-dimensional feature vector reflecting activations in the second fully-connected layer (i.e., `fc6`) of VGG-19, a higher layer in the network. We then applied dimensionality reduction (PCA) and k-means clustering on these feature vectors, yielding 70 clusters containing between 2 and 80 objects each. Among clusters that contained at least eight objects, we manually identified two visual categories containing eight objects each (Fig. 2A).

Task Procedure On each trial, both participants were shown the same set of four objects in randomized locations. One of the four objects was highlighted on the sketcher's screen to designate it as the target. Sketchers drew using their mouse cursor in black ink on a digital canvas embedded in their web browser (300×300 pixels; pen width = 5px). Each stroke was rendered on the viewer's screen in real time and sketchers could not delete previous strokes. The viewer aimed to click one of the four objects as soon as they were confident of the identity of the target, and participants received immediate feedback: the sketcher learned when and which object the viewer had clicked, and the viewer learned the true identity of the target. Both participants were incentivized to perform both quickly and accurately. They both earned an accuracy bonus for each correct response, and the sketcher was required to complete their drawings in 30 seconds or less. If the viewer responded correctly within this time limit, participants also received a speed bonus inversely proportional to the time taken until the response.

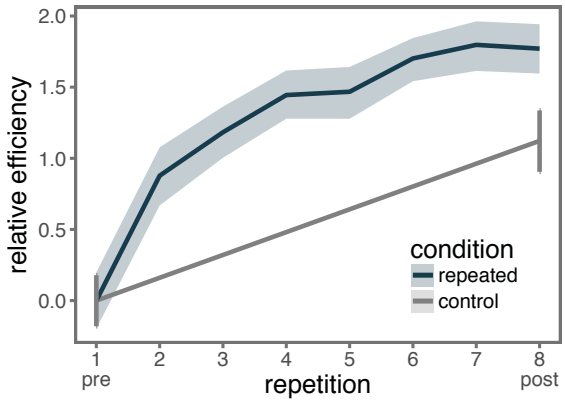

Figure 3: Communication efficiency across repetitions. Efficiency combines both speed and accuracy, and is plotted relative to the first repetition. Error ribbons represent 95% CI.

Design For each pair of participants, two sets of four objects were randomly sampled to serve as communication contexts: one was designated the *repeated* set while the other served as the *control* set (Fig. 2B).[1] The experiment consisted of three phases (Fig. 2C). During the repeated reference phase, there were six repetition blocks of four trials, and each of the four *repeated* objects appeared as the target once in each repetition block. In a pretest phase at the beginning of the experiment and a posttest phase at the end, both repeated and control objects appeared once as targets (in their respective contexts) in a randomly interleaved order.

Results

Because objects were randomly assigned to repeated and control conditions, we expected no differences in task performance in the pretest phase. We found that pairs identified the target at rates well above chance in this phase (75.7% repeated, 76.1% control, chance = 25%), suggesting that they were engaged with the task but not at ceiling performance. We found no difference in accuracy across conditions (mean difference: 0.3%, bootstrapped CI: $[-7\%, 7\%]$).

In order to measure how well pairs learned to communicate throughout the rest of their interaction, we used a measure of communicative efficiency (the *balanced integration score,* Liesefeld & Janczyk, 2018) that takes both accuracy (i.e., proportion of correct viewer responses) and response time (i.e., latency before viewer response) into account. This efficiency score is computed by first z-scoring accuracy and response time across repetitions within an interaction to map values from different interactions to the same scale, and then subtracting the standardized response time from standardized accuracy. It is highest when pairs are both fast and accurate, and lowest when they make more errors and take longer, relative to their own performance on other trials.

[1] In half of the pairs, the four control objects were from the same stimulus cluster as repeated objects; in the other half, they were from different clusters. The rationale for this was to support investigation of between-cluster generalization in future analyses. In current analyses, we collapse across these groups.

To evaluate changes in communicative efficiency, we fit a linear mixed-effects model including random intercepts, slopes, and interactions for each pair of participants. We found a main effect of increasing communicative efficiency for all targets between the *pre* and *post* phases ($b = 1.45$, $t = 14.3$, $p < 0.001$), reflecting general improvements due to task practice. Critically, however, this analysis also revealed a reliable interaction between phase and condition: communicative efficiency improved to a greater extent for repeated objects than control objects ($b = 0.648$, $t = 3.09$, $p = 0.003$; see Fig. 3). Analysis of changes in raw accuracy yielded a similar result: performance on repeated objects improved by 14.5%, while performance on control objects only improved by 7.1%. Together, these data show that there are benefits of repeatedly communicating about an object that accrue specifically to that object, suggesting the formation of object-specific graphical conventions.

Part II: What explains gains in efficiency?

Our visual communication experiment established that pairs of participants coordinate on more efficient and *object-specific* ways of depicting targets. This raises the question: to what extent do these gains in efficiency reflect the accumulation of *interaction-specific* shared knowledge between a sketcher and viewer, as opposed to the combination of task practice and the inherent visual properties of their drawings?

To disentangle the contributions of these different factors, we conducted two control experiments to estimate the how recognizable these drawings were to naive viewers outside the social context in which they were produced. Participants in one control group were shown a sequence of drawings taken from a single interaction, closely matching the experience of viewers in the communication experiment. Participants in a second control group were instead shown a sequence of drawings pieced together from many different interactions, thus disrupting the continuity experienced by viewers paired with a single sketcher. Insofar as interaction-specific shared knowledge contributed to the efficiency gains observed previously, we hypothesized that the second group would not improve as much over the course of the experimental session as the first group would.

Methods: Recognition Control Experiments

Participants We recruited 245 participants via Amazon Mechanical Turk. We excluded data from 22 participants who did not meet our inclusion criterion for accurate and consistent response on attention-check trials (see below).

Task, Design, & Procedure On each trial, participants were presented with a drawing and the same set of four objects that accompanied that drawing in the original visual communication experiment. They also received the same accuracy and speed bonuses as viewers in the communication experiment. To ensure task engagement, we included five identical attention-check trials that appeared once every eight trials. Each attention-check trial presented the same set of

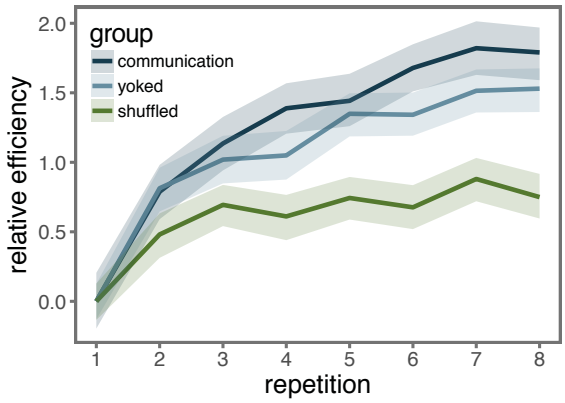

Figure 4: Comparing drawing recognition performance between viewers in communication experiment with those of yoked and shuffled control groups. Error ribbons represent 95% CI.

objects and drawing, which we identified during piloting as the most consistently and accurately recognized by naive participants. Only participants who responded correctly on at least four out of five of these trials were retained in subsequent analyses.

Each participant was randomly assigned to one of two conditions: a *yoked* group and a *shuffled* group. Each yoked participant was matched with a single interaction from the original cohort and viewed 40 drawings in the same sequence the original viewer had. Those in the shuffled group were matched with a random sample of 10 distinct interactions from the original cohort and viewed four drawings from each in turn, which appeared within the same repetition block as they had originally. For example, if a drawing was produced in the fifth repetition block in the original experiment, then it also appeared in the fifth block for shuffled participants.

At the trial level, groups in both conditions thus received exactly the same visual information and performed the task under the same incentives to respond quickly and accurately. At the repetition level, both groups received exactly the same amount of practice recognizing drawings. Thus any differences between these groups are attributable to whether drawings came from the same communicative interaction, which would support the accumulation of interaction-specific experience, or from several different interactions, where such accumulation would be minimal.

Results

Interaction-specific history enhances recognition by third-party observers We compared the yoked and shuffled groups by measuring changes in recognition performance across successive repetitions using the same efficiency metric we previously used. We estimated the magnitude of these changes by fitting a linear mixed-effects model that included group (yoked vs. shuffled), repetition number (i.e., first through eighth), and their interaction, as well as random intercepts and slopes for each participant. While we found a significant increase in recognition performance across both groups ($b = 0.18,\ t = 12.8,\ p < 0.001$), we also found a large and reliable interaction: yoked participants improved to a substantially greater degree than shuffled participants ($b = 0.10,\ t = 4.9,\ p < 0.001$; Fig. 4). Examining accuracy alone yielded similar results: the yoked group improved to a greater degree across the session (yoked: +15.8%, shuffled: +5.6%). Taken together, these results suggest that third-party observers in the yoked condition who viewed drawings from a single interaction were able to take advantage of this continuity to more accurately identify what successive drawings represented. While observers in the shuffled condition still improved over time, being deprived of this interaction continuity made it relatively more difficult to interpret later drawings.

Viewer feedback also contributes to gains in performance Unlike viewers in the interactive visual communication experiment, participants in the yoked condition made their decision based only on the whole drawing and were unable to interrupt or await additional information if they were still uncertain. Sketchers could have used this feedback to modify their drawings on subsequent repetitions. As such, comparing the yoked and original communication groups provides an estimate of the contribution of these viewer feedback channels to gains in performance (Schober & Clark, 1989). In a mixed-effects model with random intercepts, slopes, and interactions for each unique trial sequence, we found a strong main effect of repetition ($b = 0.23,\ t = 12.8,\ p < 0.001$), as well as a weaker but reliable interaction with group membership ($b = -0.05,\ t = -2.2,\ p = 0.032$, Fig. 4), showing that the yoked group improved at a more modest rate than viewers in the original communication experiment had.

To better understand this interaction, we further examined changes in the accuracy and response time components of the efficiency score. We found that while viewers in the communication experiment were more accurate than yoked participants overall (communication: 88%, yoked: 75%), *improvements* in accuracy over the course of the experiment were similar in both groups (communication: +14.5%, yoked: +15.8%). The interaction instead appeared to be driven by differential reductions in response time between the first and final repetitions (communication: 10.9s to 5.84s; yoked: 4.66s to 3.31s). These reductions were smaller in the yoked group, given that these participants did not need to wait for each stroke to appear before making a decision, and thus may have already been closer to floor.

Part III: How do visual features of drawings change over the course of an interaction?

The results so far show that repeated visual communication establishes object-specific, interaction-specific ways of efficiently referring to objects. An intriguing implication is that interacting pairs achieved this by gradually forming *ad hoc* graphical conventions about what was relevant and sufficient to include in a drawing to support rapid identification of the target object. Here we explore this possibility by examining how the drawings themselves changed throughout an interac-

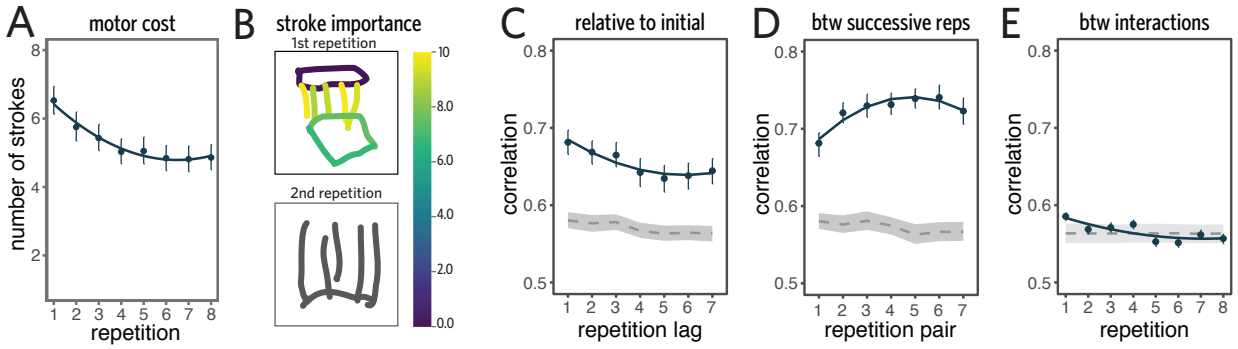

Figure 5: (A) Sketchers use fewer strokes over time. (B) Visualizing importance of individual strokes in successive drawings. (C) Drawings become increasingly dissimilar from initial drawing. (D) Drawings become more consistent from repetition to repetition. (E) The same object is drawn increasingly dissimilarly by different sketchers. Error ribbons represent 95% CI, dotted lines represent permuted baseline.

tion. Concretely, we investigated four aspects that would reflect the increasing contribution of interaction-specific shared knowledge: *first*, decreasing number of strokes used (i.e., reducing motor cost of each drawing); *second*, increasing dissimilarity from the initial drawing produced (i.e., cumulative drift from the starting point); *third*, increasing similarity between successive drawings (i.e., convergence on internally consistent ways of depicting objects within an interaction); *fourth*, increasing dissimilarity between drawings of the same object produced in different interactions (i.e., discovery of multiple viable solutions to the coordination problem).

Measuring visual similarity between drawings

Measuring visual similarity between drawings depends upon a principled approach for encoding their high-level visual properties. Here we capitalize on recent work validating the use of deep convolutional neural network models to encode such perceptual content in drawings (Fan et al., 2018). As when identifying clusters of similar object stimuli, we again used VGG-19 to extract 4096-dimensional feature vector representations for drawings of every object, in every repetition, from every interaction. Using this feature basis, we compute the similarity between any two drawings as the Pearson correlation between their feature vectors (i.e., $s_{ij} = cov(\vec{r}_i,\vec{r}_j)/\sqrt{var(\vec{r}_i)\cdot var(\vec{r}_j)}$).

Results

Fewer strokes across repetitions A straightforward explanation for the gains in communication efficiency observed in Part I is that sketchers were able to use fewer strokes per drawing to achieve the same level of viewer recognition accuracy. Indeed, we found that the number of strokes in drawings of repeated objects decreased steadily as a function of repetition in a mixed-effects model ($b = -0.216$, $t = -6.00$; Fig. 5A), suggesting that pairs were increasingly able to rely upon shared knowledge to communicate efficiently. This result raises a question about *which* strokes are preserved across successive repetitions during the formation of graphical conventions. In ongoing work, we are using a lesion method to investigate the "importance" of each stroke within a drawing for explaining similarity to the next repetition's drawing of that object. We re-render the drawing without each stroke and compute the similarity, yielding a heat map across strokes (see Fig. 5B for an example visualization). The more dissimilar the lesioned drawing without a particular stroke is to an intact version of the next repetition's drawing, the more "important" we consider that stroke to be.

Increasing dissimilarity from initial drawing Mirroring the observed reduction in the number of strokes across repetitions, we hypothesized that there was also cumulative change in the visual content of drawings across repetitions. Concretely, we predicted that drawings would become increasingly dissimilar from the initial depiction. We tested this prediction in a mixed-effects regression model including linear and quadratic terms for repetition as well as intercepts for each target and pair. We found a significant decrease in similarity to the initial round across successive repetitions, ($b = -0.62$, $t = -5.59$; Fig. 5C), suggesting that later drawings had moved to a different region of visual feature space. However, since the entire distribution of drawings may have drifted to a different region of the visual feature space for generic reasons (i.e., because they were sparser overall), we conducted a stricter permutation test. We scrambled drawings across pairs but within each repetition and target and re-ran our mixed-effects model. The observed effect fell outside this null distribution ($CI = [-3.53 - 0.88]$, $p < .001$), showing that successive drawings by the same sketcher deviated from their own initial drawing to a greater degree than would be expected due to generic differences between drawings made at different timepoints in an interaction.

Increasing internal consistency within interaction As sketchers modified their drawings across successive repetitions, we additionally hypothesized that they would converge on increasingly consistent ways of depicting each object. To test this prediction, we computed the similarity of successive drawings of the same object made in the same interaction (i.e. repetition k to $k+1$). A mixed-effects model with random intercepts for both object and pair showed that similarity between successive drawings increased substantially

throughout an interaction ($b = 0.53$, $t = 5.03$; Fig. 5). Again, we compared our empirical estimate of the magnitude of this trend to a null distribution of slope t values generated by scrambling drawings across pairs. The observed increase fell outside this null distribution ($CI = [-3.21, -0.60], p < .001$), providing evidence that increasingly consistent ways of drawing each object manifested only for series of drawings produced within the same interaction.

Increasingly different drawings across interactions Our recognition control experiments suggested that the graphical conventions discovered by different pairs were increasingly opaque to outside observers. This effect could arise if early drawings were more strongly constrained by the visual properties of a shared target object, but later drawings diverged as different pairs discovered different equilibria in the space of viable graphical conventions. Under this account, drawings of the same object from different pairs would become increasingly dissimilar from each other across repetitions. We tested this prediction by computing the mean pairwise similarity between drawings of the same object within each repetition index, but produced in different interactions. Specifically, for each object, we considered all interactions in which that object was repeatedly drawn. Then, for each repetition index, we computed the average similarity between drawings of that object. In a mixed-effects regression model including linear and quadratic terms, as well as random slopes and intercepts for object and pair, we found a small but reliable negative effect of repetition on between-interaction drawing similarity ($b = -1.4$, $t = -2.5$; Fig. 5E). We again conducted a permutation test to compare this t value with what would be expected from scrambling sketches across repetitions for each sketcher and target object. We found that the observed slope was highly unlikely under this distribution ($CI = [-0.57, 0.60]$, $p < 0.001$), even if the similarity at each round was not so unlikely.

Discussion

In this paper, we investigated the joint contributions of visual information and social context to determining the meaning of drawings. We observed in an interactive Pictionary-style communication game that pairs of participants discover increasingly sparse yet effective ways of depicting objects over repeated reference. Through a series of control experiments, we demonstrated that these conventionalized representations were both object-specific and interaction-specific: drawings were harder for independent viewers to recognize without sharing the same interaction history. Furthermore, by analyzing the high-level visual features of drawings, we found that they became increasingly consistent within an interaction, but that different pairs discover different equilibria in the space of viable graphical conventions. Taken together, our findings suggest that repeated visual communication promotes the emergence of depictions whose meanings are increasingly determined by interaction history rather than their visual properties alone.

A key experimental design choice was the use of visual objects as the targets of reference, by contrast with the verbal labels or audio clips used in prior work (Galantucci & Garrod, 2011; Fay, Garrod, Roberts, & Swoboda, 2010). As such, communication between the sketcher and viewer was grounded in the same visual information about the appearance of these objects, encouraging the production of more 'iconic' initial drawings that more strongly resembled the target object (Verhoef, Kirby, & de Boer, 2016; Perlman, Dale, & Lupyan, 2015). As their communication became increasingly efficient across repetitions, their drawings became simpler and apparently more 'abstract'. An exciting direction for future work is to develop robust and principled computational measures of the degree of visual correspondence between any drawing and any target object, thereby shedding light on the nature of visual abstraction and iconicity.

A second important design choice was the use of a speed bonus incentivizing participants to complete trials quickly. What role do such incentives play in the formation of graphical conventions? Recent computational models of visual communication have found that both how costly a drawing is to produce (i.e., time/ink) and how informative a drawing is in context are critical for explaining the way people spontaneously adjust the level of detail to include in their drawings in one-shot visual communication tasks (Fan, Hawkins, Wu, & Goodman, 2019). The consequences of this intrinsic preference for less costly drawings may be compounded across repetitions, as the accumulation of interaction history allows people to be equally informative with fewer strokes (Hawkins et al., 2017). The magnitude of these intrinsic costs may vary across individuals, however, and the speed bonus made them explicit.

A major open question raised by our work concerns how people decide what information to preserve or discard across repetitions. One possibility is that successful viewer comprehension is attributed to the most recent strokes produced, leading these to be more strongly preserved. For example, if the viewer was able to correctly identify the target only after the backrest was drawn, the sketcher may continue to selectively draw this part. Another possibility is that sketchers preserve what they judge to be the most diagnostic information about the target, regardless of when the viewer made their response. For example, sketchers may focus on drawing the backrest if it strongly distinguishes the target from distractors in context. Future work should disentangle these possibilities empirically and via development of computational models of visual communication that can learn from task-related feedback, as well as judge which strokes would be most diagnostic.

Visual communication is a powerful vehicle for the cultural transmission of knowledge. Over time, advancing our knowledge of the cognitive mechanisms underlying the formation of graphical conventions may lead to a deeper understanding of the origins of modern symbolic systems for communication and the design of better visual communication tools.

Acknowledgments

Thanks to Mike Frank and Hyo Gweon for helpful discussion. RXDH was supported by the E. K. Potter Stanford Graduate Fellowship and the National Science Foundation Graduate Research Fellowship (DGE-114747). MS was supported by the Masason Foundation Scholarship and the Center for the Study of Language and Information at Stanford.

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

References

Abell, C. (2009). Canny resemblance. *Philosophical Review*, *118*(2), 183–223.

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., ... others (2015). Shapenet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39.

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

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

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

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, *34*(3), 351–386.

Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive science*, *29*(5), 737–767.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in Human Neuroscience*, *5*, 11.

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

Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. *Psychological Review*, *103*(1), 34.

Goodman, N. (1976). *Languages of art: An approach to a theory of symbols*. Hackett.

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005–10014.

Hawkins, R. X. D. (2015). Conducting real-time multiplayer experiments on the web. *Behavior Research Methods*, *47*(4), 966-976.

Hawkins, R. X. D., Frank, M. C., & Goodman, N. D. (2017). Convention-formation in iterated reference games. In *Proc. of the 39th Annual Meeting of the Cognitive Science Society.*

Hoffmann, D., Standish, C., García-Diez, M., Pettitt, P., Milton, J., Zilhão, J., ... others (2018). U-th dating of carbonate crusts reveals neandertal origin of iberian cave art. *Science*, *359*(6378), 912–915.

Kubilius, J., Bracci, S., & de Beeck, H. P. O. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS Computational Biology*, *12*(4), e1004896.

Liesefeld, H. R., & Janczyk, M. (2018). Combining speed and accuracy to control for speed-accuracy trade-offs. *Behavior Research Methods*.

Perlman, M., Dale, R., & Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society open science*, *2*(8), 150152.

Peterson, J. C., Abbott, J. T., & Griffiths, T. L. (2018). Evaluating (and improving) the correspondence between deep neural networks and human representations. *Cognitive Science*, *42*(8), 2648–2669.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

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

Verhoef, T., Kirby, S., & de Boer, B. (2016). Iconicity and the emergence of combinatorial structure in language. *Cognitive science*, *40*(8), 1969–1994.

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*, 201403112.