Emergent Graphical Conventions in a Visual Communication Game

Shuwen Qiu[1⋆] Sirui Xie[1⋆] Lifeng Fan[3] Gao Tao[2] Song-Chun Zhu[1,2,3,4,5] Yixin Zhu[3,4]

⋆Equal contributors [1]UCLA Department of Computer Science [2]UCLA Department of Statistics
[3]Beijing Institute for General Artificial Intelligence (BIGAI) [4]Peking University [5]Tsinghua University
{s.qiu, srxie}@ucla.edu, lifengfan@bigai.ai, {tao.gao, sczhu}@stat.ucla.edu, y@bigai.ai
https://sites.google.com/view/emergent-graphical-conventions

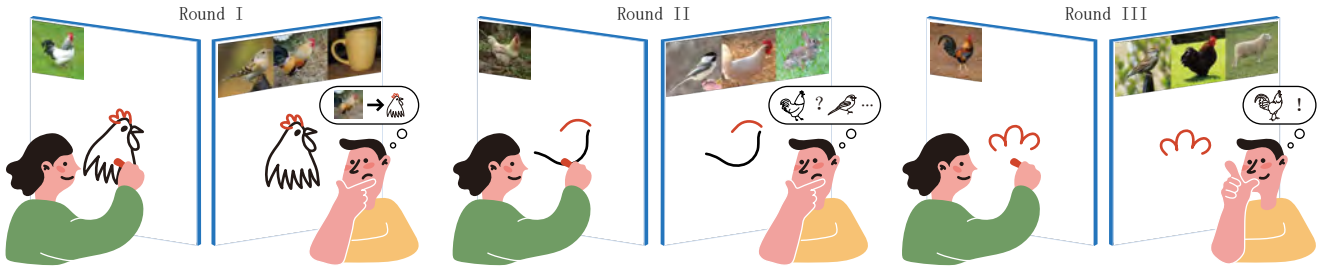


Figure 1. In an iterative sketch communication game [17], players first need to ground sketches to referents. The drawer (Alice) gradually simplifies the drawing but keeps the most salient parts of the target concept (rooster crown). This evolution process enables the viewer (Bob) to promptly distinguish the target (rooster) from distractors (bird, cup, rabbit, and sheep).

Abstract

Humans communicate with graphical sketches apart from symbolic languages [8]. While recent studies of emergent communication primarily focus on symbolic languages [24], their settings overlook the graphical sketches existing in human communication; they do not account for the evolution process through which symbolic sign systems emerge in the trade-off between iconicity and symbolicity. In this work, we take the very first step to model and simulate such an evolution process via two neural agents playing a visual communication game; the sender communicates with the receiver by sketching on a canvas. We devise a novel reinforcement learning method such that agents are evolved jointly towards successful communication and abstract graphical conventions. To inspect the emerged conventions, we carefully define three key properties – iconicity, symbolicity, and semanticity – and design evaluation methods accordingly. Our experimental results under different controls are consistent with the observation in studies of human graphical conventions [9, 17]. Of note, we find that evolved sketches can preserve the continuum of semantics [30] under proper environmental pressures. More interestingly, co-evolved agents can switch between conventionalized and iconic communication based on their familiarity with referents. We hope the present research can pave the path for studying emergent communication with the unexplored modality of sketches.

1. Introduction

Communication problem naturally arises when traveling in a foreign country where you do not speak the native language, which necessitates exploring non-linguistic means of communication, such as drawings. Due to its *iconic* nature (*i.e.*, perceptual resemblance to or natural association with the referent), drawings serve as a powerful tool to communicate concepts transcending language barriers [8]. In fact, we humans started to use drawings to convey messages dating back to 40,000–60,000 years ago [17, 18]. Some studies from cognitive science hypothesize a transition from sketch-based communication before the formation of sign systems and provide evidence that iconic signs can gradually become *symbolic* through repeated communication [8]. In contrast to *icons*, *symbols* are special forms bearing arbitrary relations to the referents. Fig. 1 describes a typical scenario of such phenomena: Alice (in green) uses a sketch to communicate the concept "rooster" to Bob (in yellow). Initially, they need to ground the sketch to the referent. Later, details of the visual concept, such as strokes of head and body, are gradually abstracted away, leaving only the most salient part, the crown. The iconicity in the communicated sketch drops while the symbolicity rises.

While models of emerging communication protocols has attracted attention [3, 4, 6, 12, 16, 24–26, 31, 33], the initial and primary communication medium is presumed and limited to be symbolic rather than iconic. By simulating a multi-agent *referential game*, prior work seeks for envi-

ronmental driving forces behind the emergence of effective communications. In a typical setup of referential games, two agents play similar roles as in the above Alice-Bob example but share a primitive set of arbitrary tokens (*i.e.*, the vocabulary). Using these tokens, an agent (the sender) attempts to communicate a message to another agent (the receiver). A communication convention is emerged when two agents successfully communicate by associating visual concepts in the images with tokens in the pre-selected vocabulary. Though this line of work has probed into some linguistic properties of the established communication conventions [25, 26, 33], some intriguing questions remain open: How do agents make a trade-off between iconicity and symbolicity to emerge symbolic sign systems?

In this work, we present the very first step of modeling and simulating the evolution process of *graphical conventions* [17], a two-participant communication convention whose medium is drawings in an abstract form. Specifically, we consider our contributions in three folds:

First, we model a multi-agent visual communication game and propose a learning framework, wherein the sender and the receiver evolves jointly. This visual communication game is an alternating sequential decision-making process, in which the sender generates a sequence of strokes step by step, terminated by the receiver. In contrast to discretized tokens in prior work, strokes can be naturally parametrized in a continuous space [13, 19] such that the derivatives of learning objectives can be more effectively back-propagated through communication channels [11]. We further derive a novel training surrogate for multi-agent reinforcement learning based on a joint value function and the eligibility trace method [39]. In experiments, we empirically demonstrate that such an integration of function approximation and Monte Carlo sampling in sequential communication facilitates the agents to be aware of the correlation between complex and simple sketches, hereby enabling a smooth abstraction process.

Second, we define essential properties in studying evolved sketches. Specifically, we define *iconicity* [8] as the drawings exhibiting high visual resemblance to the corresponding images, such that they are proximal to the latter when measured on the high-level embedding of a general-purpose visual system; we define *symbolicity* [10] as these drawings being consistently separable in the high-level visual embedding, which facilitates new communication participants to easily distinguish between categories without grounding them to referents; we define *semanticity* [15] as the topography of the high-level embedding space of the drawings being strongly correlated to that of images, such that semantically similar instances and categories lie close to each other in the embedding space. Of note, this is not the only way to define these cognitive concepts; our intention is to align readers on critical concepts in our work.

Third, we present a suite of quantitative and qualitative methods to evaluate the emergent graphical conventions based on the carefully defined *iconicity*, *symbolicity*, and *semanticity*. This is neccesary because a high communication rate does not imply good representations [2]. The graphical nature of the communication medium mandates us to repurpose representation learning metrics rather than adopt linguistic metrics in emergent symbolic communication. We evaluate the contribution of different environmental drivers, early decision, sender's update, and sequential communication, to the three properties of the emergent conventions. Critically, the empirical results assessed on our metrics well align with our prediction based on the findings of human graphical conventions [9, 17], justifying our environment, model, and evaluation. One of these setups emerges conventions where the three properties are consistent with our expectation of a sign system. Particularly, we find two inspiring phenomena: (i) Evolved sketches from semantically similar classes are perceptually more similar to each other than those falling into different superclasses. (ii) To communicate concepts not included in their established conventions, evolved agents can return to more iconic communications as we humans do. We hope our work can invoke the investigation of emergent communication in the unexplored modality of sketches and facilitate the study of cognitive evolutionary theories of pictographic sign systems.

2. Related work

Learning to sketch Ha and Eck [13] begin the endeavor of teaching modern neural models to sketch stroke by stroke. However, generating meaningful stroke sequences directly from various categories of natural images is still in early phase [38, 42, 48]. To assure the interestingness of the category-level sketch communication, we design a stage-wise agent that first transfers a natural image into a pixel-level sketch through a CNN-based model [20] and then draws the sketch on the canvas stroke by stroke with a policy [19]. To pretrain our neural agents to sketch, we select Sketchy [34] from many datasets [5, 13, 34, 45] for its fine-grained photo-sketch correspondence, rich stroke-level annotations, and well-organized categorization structure.

Communication games While learning to sketch is formulated as a single-agent task with explicit supervision, our focus is on **how sketches would evolve** when utilized as the communication medium between two cooperative agents. The tasks the two agents cooperate on are always formulated as communication games, recently adopted to study phenomena in natural languages, such as symbolic language acquisition [12] and the emergence of compositionality [33]. Some notable works [6, 16, 25, 26] have devised interesting metrics, such as *purity* [26] and *topographic similarity* [25]. In comparison, our work is unique

due to the distinctive communication medium, continuously parametrized sketches. Although a concurrent work [29] also enables the agents to sketch in a communication game, it focuses only on drawing interpretable sketches without abstracting into graphical symbols along the communication process. We position our work as an alternative to emergent symbolic communication, since the emergent graphical symbols may better represent the continuum of semantics, as encoded in the vector representation of tokens [30]. Methodologically, we devise new evaluation metrics for this unexplored modality, assessing the *iconicity*, *symbolicity*, and *semanticity* in the evolved sketches.

Emergent graphical conventions Evolving from sketches to graphical conventions/symbols is an active field in cognitive science under the banner of "emergent sign systems." Fay *et al*. [9] show that pair interaction can form their local conventions when they play the Pictionary game. Using natural images instead of texts as the prompt, Hawkins *et al*. [17] show that, besides partners' shared interaction history, visual properties of the images also influence the formed graphical conventions. That is, the evolved sketches highlight visually salient parts. Nevertheless, only a few computational models exist apart from these human behavior studies. Fan *et al*. [7] describe a model for selecting complex or simple sketches considering the communication context. Bhunia *et al*. [1] and Muhammad *et al*. [32] consider stroke selection and reordering to simplify the sketch. In contrast to sketch or stroke selection, we model **embodied** agents who can draw and recognize sketches, a more natural setting if we were to fulfill the goal of modeling the transition from iconic sketches to graphical symbols.

3. The visual communication game

Our visual communication game is formulated as a tuple

$$(\mathcal{I}, C, \mathcal{A}_S, \mathcal{A}_R, G, r, \gamma),$$

where $\mathcal{I}$ is the image set to be presented to the sender S and the receiver R. These images contain a single object in the foreground, and hence the image space $\mathcal{I}$ can be partitioned into N classes according to the category of the objects. In each round of the game, the sender is presented with one image I_S, and the receiver is presented with M images $\{I_R^1, ..., I_R^M\}$. Inspired by the category-level game introduced by Lazaridou *et al*. [26], we make the observations of S and R disjoint (*i.e*., $I_S \notin \{I_R^1, ..., I_R^M\}$), but with a target image I_R^* in the same class as I_S. We refer to the M images that the receiver can see as the *context*. Neither the receiver nor the sender would see the image(s) presented to their partner; they can only communicate this information by drawing sketches on the canvas C, observable to both players. As shown in Fig. 2, at the beginning of each round, C_0 is initialized to be blank. Only the sender can draw on the canvas with actions chosen from $\mathcal{A}_S$. The action at each time step consists of 5 strokes, which are continuous vectors in $\mathbb{R}^6$. We constrain each dimension to be in $(0, 1)$ due to limited space on the canvas. The canvas is updated from C_t to C_{t+1} by the renderer G after each step of the sender's sketching. The receiver, after observing the updated canvas, would have to choose among the M images or wait for the next step from the sender; these $M+1$ possible actions constitute $\mathcal{A}_R$. A game round terminates when the receiver gives up waiting and chooses one from the images. After the termination, the sender and the receiver will receive a shared reward or penalty, depending on if the receiver makes the right choice:

$$r : \mathcal{I} \times \mathcal{I} \to \{-1, 1\}.$$

This reward/penalty is temporally decayed with a decay factor γ. That is, if the receiver decides to choose from the images at step t, this cooperating pair will receive either γ^t or $-\gamma^t$. Hence, even though the players do not receive an explicit penalty for long conversations, there is an implicit penalty/reward for delayed positive/negative answers. No reward will be assigned if the receiver chooses to wait. The next round starts after the reward/penalty assignment.

4. Agents

The two agents involved in the visual communication game are modeled with two decision policies, π_S and π_R, for the sender and the receiver, respectively. These policies are stochastic mapping from the agents' observation space to the action space:

$$\pi_S : \mathcal{I} \times C \to \mathcal{P}(\mathcal{A}_S), \quad \pi_R : \mathcal{I}^M \times C \to \mathcal{P}(\mathcal{A}_R), \qquad (1)$$

where $\mathcal{P}(\mathcal{A})$ is a distribution over the support set $\mathcal{A}$. As shown in Fig. 2, at each time step $t \in \{0, ...T\}$, the sender first emits the stroke parameters for the next five strokes $a_{St} \sim \pi_S(I_S, C_t)$. These strokes are applied to the canvas by a differentiable renderer, $C_{t+1} = G(C_t, a_{St})$. Next, the updated canvas is transmitted to the receiver. The receiver decides whether to terminate the game and make its prediction (*i.e*., $a_{Rt} \in \{1, ..., M\}$) or wait for another round (*i.e*., $a_{Rt} = M+1$); its decision is sampled from π_R. If a prediction is made, it is used to select the image $I_R^{a_R}$ from $\{I_R^1, ...I_R^M\}$ and score this game with $r(I_S, I_R^{a_R})$. Otherwise, this routine repeats in the next step $t \leftarrow t+1$.

4.1. Sender

Prior to playing the visual communication game, the sender should be able to (i) extract edges from natural images [44] and (ii) draw sketches that closely resemble the configurations of salient edges [28], just as humans [35]

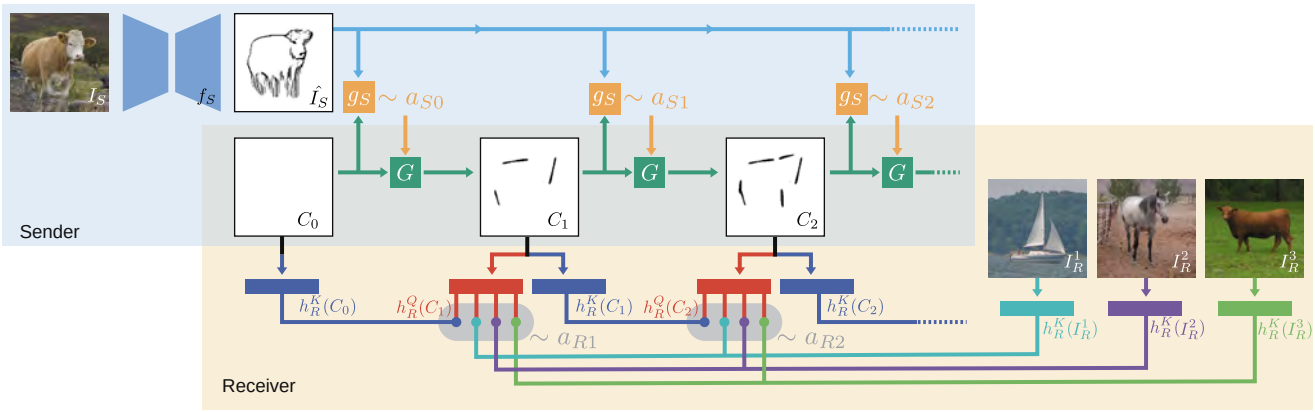


Figure 2. **Communication process**. In our visual communication game, a sender S and a receiver R only share the observation of the canvas C. The sender first converts the natural image I_S to a pixel-level sketch $\hat{I}_S$. At each step, the sender first draws five strokes a_S through the renderer G, which updates the canvas to C_{t+1}. Next, the receiver uses the updated canvas C_{t+1} to query from the context images $\{I_R^1, ..., I_R^M\}$ and the last canvas C_t, deciding the action a_R at this step. The game continues if the receiver chooses to wait.

do. To endow the sender with these capabilities, we design a stage-wise architecture $h_S = g_S \circ f_S$. Specifically, I_S is first converted to a target sketch $\hat{I}_S$ using a visual module f_S [20], capturing the salient edge information in the natural image; we directly adopt the pretrained model from the referred work. Next, $\hat{I}_S$ is concatenated with the current canvas C_t and fed to the sketching module g_S, whose architecture is built upon Huang *et al*. [19]. This sketching module outputs five vectors in the form $(x_0, y_0, x_1, y_1, x_2, y_2)$, which parametrizes the curve of one stroke. The policy is parametrized as a Gaussian distribution during training,

$$\pi_S = \mathcal{N}(\mu_t, \sigma^2), \quad \mu_t = h_S(I_S, C_t), \quad \sigma^2 = c \cdot \mathbf{I}, \tag{2}$$

where $\mathbf{I}$ is the identity matrix, and c is a constant hyperparameter. During testing, we set $c = 0$.

These stroke parameters a_{St} are fed into a pretrained renderer G [19] to update the canvas, $C_{t+1} = G(C_t, a_{St})$. This renderer is fully differentiable, enabling end-to-end model-based training [14] of the sketching module g_S. We pretrain g_S on Sketchy [34]; see Supp. for results.

4.2. Receiver

The receiver, similar to the sender, should also carry some rudimentary visual capability to this game. Unlike the low-level vision needed for the sender, the requirement for the recevier is high-level visual recognition. Therefore, we adopt a pretrained VGG16 [37] as the visual module $f_R : \mathcal{I} \to \mathbb{R}^{4096}$ of the receiver, following a similar practice in recent literature [16, 26]. The output of this visual module is a vector, and further transformed by two separate linear layers, g_R^K and g_R^Q, into visual embeddings, $h_R^K(I)$ and $h_R^Q(I)$. That is, $h_R^K = g_R^K \circ f_R$, $h_R^Q = g_R^Q \circ f_R$.

When observing both the context $\{I_R^1, ..., I_R^M\}$ and the canvas C_t, the receiver first embeds each of them with h_R. Next, it makes the decision based on the similarity between the current canvas and each option in the context. The decision module is thus realized by a Boltzmann distribution based on resemblance:

$$\pi_R(a_{Rt} | I_R^1, ...I_R^M, C_{t-1}, C_t) = \frac{\exp(h_R^Q(C_t) \cdot h_R^K(I_R^{a_{Rt}}))}{\sum_{m=1}^{M+1} \exp(h_R^Q(C_t) \cdot h_R^K(I_R^m))}, \tag{3}$$

where $I_R^{M+1} = C_{t-1}$. Although a similar policy was proposed before [16, 25], our π_R is distinct as it is endowed with an external memory of C_{t-1}. Intuitively, if the receiver finds the current canvas C_t closer to the last canvas C_{t-1} in the embedding space than all M options in the context, it will choose to emit $a_{Rt} = M + 1$ and delay the decision to the next step; a prediction can only be made when the receiver finds the current canvas is informative enough. As a result, the sender would draw informative strokes as early as possible to avoid the implicit penalty in the decayed reward.

4.3. Learning to communicate

Policies of the sender and the receiver are trained jointly to maximize the objective of the stochastic game in Sec. 3:

$$\pi_S^*, \pi_R^* = \arg\max_{\pi_S, \pi_R} \mathbb{E}_{\tau \sim (\pi_S, \pi_R)} [\sum\nolimits_t \gamma^t r_t], \tag{4}$$

where $\tau = \{C_0, a_{S0}, C_1, a_{R1}, a_{S1}, ...\}$ is the simulated episodic trajectory. As well known in reinforcement learning, the analytical expectation in Eq. (4) is intractable to calculate along the trajectory τ. We devise value functions $\mathcal{V}(X_t)$ and $V_\lambda(X_t)$ for an optimization surrogate:

$$\mathcal{V}(X_t) = \mathbb{E}_{\pi_S(a_{St} | I_S, C_{t-1}), \pi_R(a_{Rt} | \hat{X}_t)} [(r_t + \gamma \delta(a_{Rt}) V_\lambda(X_{t+1})], \tag{5}$$

where $\hat{X}_t = [I_R^1, ...I_R^M, C_{t-1}, C_t]$, $X_t = \text{cat}([I_S], \hat{X}_t)$, $\delta(a_{Rt})$ is the Dirac delta function that returns 1 when

the action is *wait* and 0 otherwise. The expectation $\mathbb{E}_{\pi_S(a_{St}|I_S,C_{t-1})}[\cdot]$ is approximated with the point estimate, as in the reparametrization in VAE [23]. The expectation $\mathbb{E}_{\pi_R(a_{Rt}|\hat{X}_t)}[\cdot]$ can be calculated analytically because π_{Rt} is a categorical distribution. The connection between these two expectation is one of our contributions. Of note, C_t in $\hat{X}_t$ in $\pi_{Rt}(a_{Rt}|\hat{X}_t)$ is generated by the differentiable renderer G with the actions a_{St} from the sender policy $\pi_S(a_{St}|I_S,C_{t-1})$. Hence, we can have both $\partial\mathcal{V}/\partial\pi_{Rt}$ and $\partial\mathcal{V}/\partial\pi_{St}$ based on Eq. (5). This results in a novel multi-agent variant of the general policy gradient [36, 40].

$V_\lambda(X_t)$ in Eq. (5) is an eligibility trace approximation [39] of the ground-truth value function. Intuitively, a value estimate with eligibility trace V_λ mixes the bootstrapped Monte Carlo estimate $V_N^k(X_t) = \mathbb{E}_{\pi_S,\pi_R}[\sum_{n=t}^{h-1}\gamma^{n-t}r_n + \gamma^{h-t}\delta(a_{Rh})v_\phi(X_h)]$ at different roll-out lengths k, with $h = min(t+k, T_{choice})$ being the maximal timestep. T_{choice} is the timestamp when the receiver stops waiting. The articulation of such termination also makes our eligibility trace deviate from the general derivation with infinite horizon. We derive an episodic version as

$$V_\lambda(X_t) = \begin{cases} (1-\lambda)\sum_{n=1}^{H-1}\lambda^{n-1}V_N^n(X_t) + \lambda^{H-1}V_N^H(X_t) \\ \qquad\qquad \text{if } t \leqslant T_{choice} \\ v_\phi(X_t) \qquad \text{otherwise} \end{cases} \tag{6}$$

where $H = T_{choice} - t + 1$. Please refer to the supplementary material for a detailed derivation and the algorithm. Finally, $v_\phi(X_t)$ is trained by regressing the value returns:

$$\phi^* = \arg\max_\phi \mathbb{E}_{\pi_S,\pi_R}[\sum\nolimits_t \frac{1}{2}||v_\phi(X_t) - V_\lambda(X_t)||^2]. \tag{7}$$

5. Experiments

5.1. Settings

Images We used the Sketchy dataset [34] as the source of images. Due to the limited performance of the sketching module on open-domain image-to-sketch sequential generation, we select 40 categories (10 images per category) that enable satisfactory sketching behaviors.

Environmental drivers With the visual communication game and the learning agents at hand, we investigate the causal factors in emergent graphical conventions with controlled experiments. Tab. 1 lists all designed settings. Specifically, we consider the following factors:

- *Can receiver make early decisions?* The hypothesis is that the receiver's decision before exhausting the allowed communication steps may inform the sender the marginal benefit of each stroke and incentivize it to prioritize the most informative strokes. The corresponding control setting is *max-step*, where the receiver can only make the choice after the sender finishes its drawing at the maximum step. This factor is on in other settings.
- *Can sender change its way of drawing?* The hypothesis is that the mutual adaptation of the sender and the receiver may lead to better abstraction in the evolved sketches. Particularly, the sender may develop new ways of drawing in the evolution. The corresponding control setting is *sender-fixed*, wherein we freeze the parameters of the sender, such that the receiver has to adapt to its partner. This factor is on in other settings.
- *Is the game sequential, and can the receiver observe more complex drawings?* The hypothesis is that the modeling of a sequential decision-making game and the evolution from more complex sketches may regularize the arbitrariness, which is unique for graphical conventions. The corresponding control setting is *one-step*: There only exists one step of sketching, thus no sequential decision-making at all. This factor is on in other settings.

Training, validation and generalization test We train the sender and the receiver with batched forward and back-propagation, with a batch size of 64 and maximum roll-out step $T = 7$. We update using Adam [22] with the learning rate 0.0001 for a total of 30k iterations. Except for the *sender-fixed* setting, there is a warm-up phase in the first 2000 iterations for the sender where we linearly increase the learning rate to 0.0001. After the warm-up phase, the learning rate of both agents will be decreased exponentially by $0.99^{\frac{i-2000}{1000}}$, where i is the number of training iterations. In all settings, we set $M = 3$, $\gamma = 0.85$. Between iterations, we randomly sample another 10 batches for validation. We held-out 10 images per category for unseen-instance test and 10 categories for unseen-class test. Each image will be communicated as the target, resulting in 300+100 pairs in the test set. Results below are statistics of 5 random seeds.

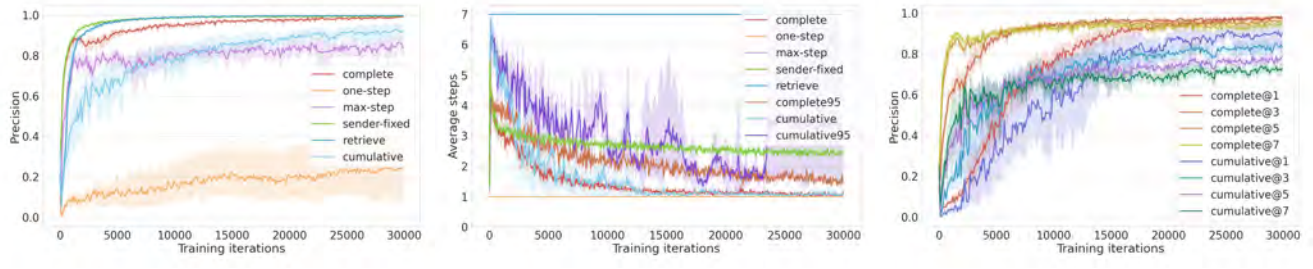
5.2. Results

5.2.1 Communication efficacy and sketch abstraction

We record both the success rate and the communication steps over the training iterations; see Fig. 3. In Fig. 3a, agents in all settings except *one-step* converge to a success rate above 80%. Among them, the communicating pairs in the *complete* setting and the *sender-fixed* setting evolve to achieve a comparable success rate with the *retrieve* baseline. Interestingly, these two pairs also emerge a phenomenon resembling the natural observation in human study, named *systematic reduction* [27]: The average steps first increase and then gradually decrease as in Fig. 3b. Contrasting *complete* and *sender-fixed*, we can see: (i) The emergent conventions in the former is much simpler than the latter (less steps in Fig. 3b), which implies the contribution of mutual adaptation in sketch abstraction. (ii) The

Game Settings					Communication Accuracy (%) ± SD (avg. step)		
early decision	update sender	max/one step	description	setting names	seen	unseen instance	unseen class
yes	yes	max	our experimental setting	complete	98.07 ± 0.01(1.03)	70.37 ± 0.04(2.36)	39.40 ± 0.05(3.76)
no	yes	max	control setting for early decision	max-step	86.27 ± 0.03(7.00)	67.93 ± 0.02(7.00)	38.40 ± 0.04(7.00)
yes	no	max	control setting for evolving sender	sender-fixed	99.60 ± 0.01(2.41)	71.80 ± 0.02(3.83)	45.40 ± 0.02(4.75)
yes	yes	one	control setting for sequential game	one-step	22.87 ± 0.23(1.00)	14.07 ± 0.15(1.00)	9.60 ± 0.09(1.00)
no	no	max	baseline for all settings above	retrieve	99.47 ± 0.01(7.00)	76.80 ± 0.02(7.00)	48.00 ± 0.02(7.00)

Table 1. **Game settings and results.** The first three columns represent the configurations of the environmental drivers. Setting names and descriptions specify our purposes for intervention. The last three columns show success rates and conversation length in testing games. "seen" are validation games with the same image set as training. "unseen" are testing games with unseen images.



(a) validation accuracy (b) average communication steps (c) prediction accuracy

Figure 3. **Statistics aggregated over all random seeds.** (a) The validation accuracy of different game settings and the ablation baseline. (b) The average communication steps under different settings and ablation baselines. γ is 0.85 by default, 0.95 in complete95 and cumulative95. (c) The prediction accuracy when receivers are presented with sketches drawn by corresponding senders at 1, 3, 5, and 7 time steps, respectively. These sketches are collected in a standalone roll-out after each iteration, where early decision is disabled; agents are trained with the *complete* setting, where ealy decision is enabled.

success rate in Fig. 3a in the former converges a bit more slowly, which is reasonable since the senders can explore to change the way of drawing. In comparison, if the receiver cannot make early decisions, it has no intention to relate sketches (*i.e.*, C_{t-1} and C_t) at consecutive timesteps. The sender is thus *unaware* of each stroke's marginal information gain, which in return makes their learning harder. This might explain the relatively low success rate in the *max-step* setting. The failure of the *one-step* pairs reveals the irreplacable roles of sequential decision-making and observing complex sketches in emergent graphical communication.

To further inspect how our proposed modeling and learning on sequential decision-making facilitate the desired evolution in the sketches, we conduct an ablation study by comparing our proposed learning surrogate (Eq. (5)) and a vanilla policy gradient baseline, REINFORCE [43] with Monte Carlo cumulative rewards $\mathbb{E}_{\pi_S,\pi_R}[\sum_t[\nabla \log \pi_R(a_{Rt}|\hat{X}_t)\sum_{n=t}^{T}\gamma^{n-t}r_n]]$.

Our comparison spans three axes. First, the REINFORCE baseline converges much more slowly than the proposed surrogate; see *cumulative* vs *complete* in Fig. 3a. Second, we check the robustness under variation of decay factor γ. As shown in Fig. 3b, while the proposed method shows stable convergence in the communication steps under $\gamma = 0.85$ and $\gamma = 0.95$, the REINFORCE baseline exhibits high-variance behavior under certain setup ($\gamma = 0.95$). Finally, we check if agents' early terminations are *caused* by their *awareness* of the indistinguishable performance in longer and shorter episodes. Given a *precondition* that the longer the episodes are, the earlier the success rate increases, it should be the increase in the average performance of shorter episodes that *causes* the average timesteps to decrease. Taking 1-step and 3-step communication for example, in the *complete* setting, we shall see the success rate of the 3-step to achieve high earlier, which is then caught up but not exceeded by the 1-step. The not exceeding condition is a crucial cue to validate that the communicating partners were *actively* pursuing the Pareto front [21] / efficiency bound [46] of accuracy and complexity. This is exactly what our proposed method emerges as shown in Fig. 3c. In contrast, in the REINFORCE baseline, under the same decay factor, the performance of 1-step surpasses 3-step communication. It seems as if the incapability of *learning* long episodes *caused* the agents to *avoid* them.

All our results on success rate and communication steps are consistent with predictions based on our hypotheses, which justifies our environments and models. However, a high success rate does not necessarily imply high convention quality. Another essential property for conventions is *stability* [27]: There should exist common patterns in the re-

peated usage of conventions to convey similar concepts. We take the viewpoint of representation learning and concretize the vague *stability* with three formally defined properties: *iconicity*, *symbolicity* and *semanticity*. In the following, we introduce our experiments to measure these properties.

5.2.2 Iconicity: generalizing to unseen images

We start from *iconicity* since it is the most distinctive property in visual communication. We define *iconicity* as the drawings exhibiting high visual resemblance with the corresponding images, such that they are proximal to the latter when measured on the high-level embedding of a general-purpose visual system. Based on this definition, a more *iconic* drawing should facilitate more *generalizable* graphical conventions. Namely, in more naturalistic open-domain communication, agents would always see novel scenes containing known or unknown concepts. They should still be able to communicate with established conventions or with unconventionalized iconic drawings. Such *generalizability* can be measured in two test cases: unseen instances of seen classes and unseen classes. A successful generalization to unseen instances implies senders' ways of drawing preserve *iconicity* in the evolution. A successful generalization to unseen classes (*i.e.*, zero-shot generalization) is more difficult than unseen instances; hence, partners may increase the conversation length and communicate with more complex sketches. This requires both the senders to preserve *iconicity* in drawings and the receivers to be sensitive to the information change in the communicated sketches.

Tab. 1 reports the success rates and average timesteps in our generalization tests. The *retrieve* setting is the baseline, since there is no evolution at all and the sketches should resemble the original images the most (*i.e.*, possessing the highest *iconicity*). Unsurprisingly, its generalization performance upper-bounds all other settings. Among the experimental and controlled settings, the *complete*, the *max-step*, and the *sender-fixed* agents generalize relatively well in unseen instances (70.37 ± 0.04, 67.93 ± 0.02, 71.80 ± 0.02) and generalize above chance ($39.40 \pm 0.05, 38.40 \pm 0.04, 45.40 \pm 0.02 > 25.00$) in unseen classes. Interestingly, *complete* and *sender-fixed* communicating partners intelligently turn to longer episodes for better generalization, better than the *max-step* agents. This finding implies the partners may turn to more iconic communication when there is no established conventions/symbols, just as we humans do. Strikingly, the *max-step* conventions seem to loose more *iconicity*, possibly due to confusion on marginal information gains. The *one-step* drawings seem to lack *iconicity*.

5.2.3 Symbolicity: separating evolved sketches

Next, we measure *symbolicity* to evaluate the graphical conventionalization. We define *symbolicity* as the drawings being consistently separable in the high-level visual embedding, which facilitates new communication partners to easily distinguish between categories without grounding them to referents. Based on this definition, a more *symbolic* drawing should be more *easily separable* into their corresponding categories. To measure such *separability*, we use a pre-trained VGG16 as the new learner and finetune the last fully connected layer to classify evolved sketches into the 30 categories. Technically, we first get the 300 final canvases from the communication game, 10 for each category. Among them, we use 70% for training and 30% for testing.

Figure 4. **Testing results of classifiers trained with sketches from different settings.** *img* denotes images, and *retrieve* denotes unevolved sketches.

The bar plot in Fig. 4 shows the classification results. Since agents in the *one-step* setting do not play the game successfully, they may not form a consistent way to communicate. Agents in the *complete* setting achieve the highest accuracy, higher even compared with the result of the original images. This finding indicates that agents in the *complete* setting agree on a graphical convention that consistently highlights some features across all training instances in each category, which are also distinguishable between categories. Comparing the *max-step* with the *complete* setting, we know that early decision is an important factor for more *symbolic* conventions. Comparison between the *sender-fixed* setting and the *complete* setting suggests that the sender's evolution also contributes to high *symbolicity*.

5.2.4 Semanticity: correlating category embedding

The last desired property of graphical conventions is that the evolved sketches can preserve the perceptual metric in images [47]. We define *semanticity* as the topography of the high-level visual embedding space of drawings being strongly correlated to that of images, such that semantically similar instances and categories lie close to each other in the embedding space. To examine such *topographic correlation*, we project the embeddings obtained in Sec. 5.2.3 to a 2D space via t-SNE [41]. Fig. 5 show the visualization of the original images and the sketches in the *retrieve* and *complete* settings; please refer to supplementary for results of other settings. Images/drawings from the same category are marked in the same color. As shown, boundaries between categories are clearer in the evolved drawings in

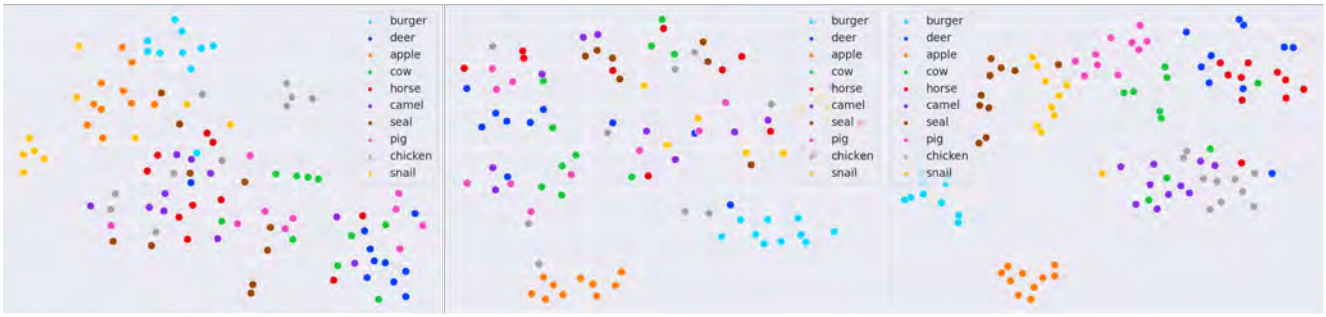


Figure 5. **t-SNE of visual embedding of the original images (left), unevolved sketches in the *retrieve* setting (middle), and evolved sketches in the *complete* setting (right).** These embeddings are from the finetuned VGGNets in Sec. 5.2.3. The evolved sketches have clearer boundaries due to the discrimination game. But they still maintain the topography that similar concepts are close to each other.

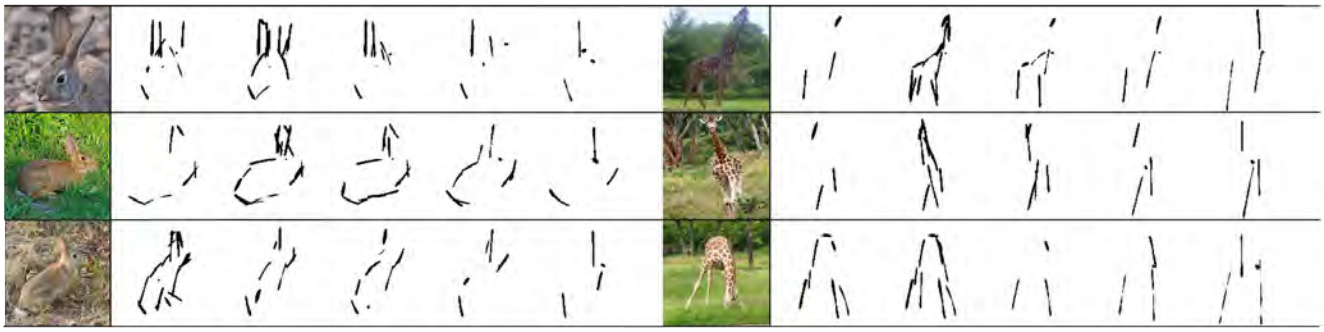


Figure 6. **Sketch evolution of rabbit and giraffe through game iterations.** For each example, sketches from the left to the right show the change of the final-step canvas from iteration 0 to 30,000. Please refer to Supp for more results.

the *complete* setting than the unevolved sketches in *retrieve* or original images; but semantically similar categories are still close to each other. For example, cow, deer, horse, and camel are proximal to each other, while burger and apple are far from them. These results highlight another uniqueness of visual communication over its symbolic counterpart: The similarity in the visual cues in the conventions may hint the *semantic* correlation between the referents.

5.2.5 Visualizing sketch evolution

To better understand the nature of the emerged conventions, we inspect the intermediate sketches in the evolution processes. Specifically, we choose to visualize the process under the *complete* setting. Fig. 6 shows three instances in two categories. For each example, drawings from the left to the right show the change of the final-step canvas from iteration 0 to 30,000. Sketches' complexity gradually decreases after an initial increase, echoing the trend of reduction described in Sec. 5.2.1. For rabbit, at the beginning, the strokes may depict instances from different perspectives; through iterations, they converge to highlight the rabbit's long ear. As for giraffe, the agents gradually learn to emphasize the long neck. Particularly, in the third example, although the giraffe lowers its head, we can still see an exaggerated vertical stroke for the neck, similar to the first example where the giraffe's head is raised. These examples show how the sender gradually unifies drawings of the same category. It can also be seen that after evolution, the sender is inclined to use the first several strokes to depict the most salient parts of visual concepts.

6. Conclusion

In this work, we present the first step of modeling and simulating the evolution of graphical conventions between two agents in a visual communication game. Agents modeled in the proposed framework can successfully communicate visual concepts using sketches as the medium. We measure the emergent graphical conventions over three carefully defined properties, *iconicity*, *symbolicty*, and *semanticity*. The experimental results under different controls suggest that early decision, mutual adaptation, and sequential decision-making can encourage *symbolicity* while preserving *iconicity* and *semanticity*. However, there are some limitations regarding the two-stage pretrained senders. An ideal sender would not need to convert the images to a sketches before it starts sketching. The limitations in the pretrained sketching module also constrain the discriminative need among the selected classes in our game. We will investigate and resolve these limitations in future research. With the uniqueness of visual conventions demonstrated, we hope our work can invoke the study of emergent communication in the unexplored modality of sketches.

References

[1] Ayan Kumar Bhunia, Ayan Das, Umar Riaz Muhammad, Yongxin Yang, Timothy M Hospedales, Tao Xiang, Yulia Gryaditskaya, and Yi-Zhe Song. Pixelor: A competitive sketching ai agent. so you think you can sketch? *ACM Transactions on Graphics (TOG)*, 39(6):1–15, 2020. 3

[2] Diane Bouchacourt and Marco Baroni. How agents see things: On visual representations in an emergent language game. In *Proceedings of the conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018. 2

[3] Kris Cao, Angeliki Lazaridou, Marc Lanctot, Joel Z Leibo, Karl Tuyls, and Stephen Clark. Emergent communication through negotiation. In *International Conference on Learning Representations (ICLR)*, 2018. 1

[4] Tom Eccles, Yoram Bachrach, Guy Lever, Angeliki Lazaridou, and Thore Graepel. Biases for emergent communication in multi-agent reinforcement learning. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2019. 1

[5] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? *ACM Transactions on Graphics (TOG)*, 31(4):1–10, 2012. 2

[6] Katrina Evtimova, Andrew Drozdov, Douwe Kiela, and Kyunghyun Cho. Emergent communication in a multi-modal, multi-step referential game. In *International Conference on Learning Representations (ICLR)*, 2018. 1, 2

[7] Judith E Fan, Robert D Hawkins, Mike Wu, and Noah D Goodman. Pragmatic inference and visual abstraction enable contextual flexibility during visual communication. *Computational Brain & Behavior*, 3(1):86–101, 2020. 3

[8] Nicolas Fay, Mark Ellison, and Simon Garrod. Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition*, 22(2):244–263, 2014. 1, 2

[9] Nicolas Fay, Simon Garrod, Leo Roberts, and Nik Swoboda. The interactive evolution of human communication systems. *Cognitive Science*, 34(3):351–386, 2010. 1, 2, 3

[10] Nicolas Fay, Bradley Walker, Nik Swoboda, and Simon Garrod. How to create shared symbols. *Cognitive Science*, 42:241–269, 2018. 2

[11] Jakob N Foerster, Yannis M Assael, Nando De Freitas, and Shimon Whiteson. Learning to communicate with deep multi-agent reinforcement learning. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2016. 2

[12] Laura Graesser, Kyunghyun Cho, and Douwe Kiela. Emergent linguistic phenomena in multi-agent communication games. In *Proceedings of the conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019. 1, 2

[13] David Ha and Douglas Eck. A neural representation of sketch drawings. In *International Conference on Learning Representations (ICLR)*, 2018. 2

[14] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations (ICLR)*, 2019. 4

[15] Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain. Semantic similarity from natural language and ontology analysis. *Synthesis Lectures on Human Language Technologies*, 8(1):1–254, 2015. 2

[16] Serhii Havrylov and Ivan Titov. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2017. 1, 2, 4

[17] Robert XD Hawkins, Megumi Sano, Noah D Goodman, and Judith W Fan. Disentangling contributions of visual information and interaction history in the formation of graphical conventions. In *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci)*, 2019. 1, 2, 3

[18] Dirk L Hoffmann, Chris D Standish, Marcos García-Diez, Paul B Pettitt, James A Milton, João Zilhão, José Javier Alcolea-González, Pedro Cantalejo-Duarte, Hipólito Collado, Rodrigo De Balbín, et al. U-th dating of carbonate crusts reveals neandertal origin of iberian cave art. *Science*, 359(6378):912–915, 2018. 1

[19] Zhewei Huang, Wen Heng, and Shuchang Zhou. Learning to paint with model-based deep reinforcement learning. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2019. 2, 4

[20] Moritz Kampelmuhler and Axel Pinz. Synthesizing human-like sketches from natural images using a conditional convolutional decoder. In *Proceedings of Winter Conference on Applications of Computer Vision (WACV)*, 2020. 2, 4

[21] Il Yong Kim and Oliver L De Weck. Adaptive weighted-sum method for bi-objective optimization: Pareto front generation. *Structural and multidisciplinary optimization*, 29(2):149–158, 2005. 6

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 5

[23] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 5

[24] Angeliki Lazaridou and Marco Baroni. Emergent multi-agent communication in the deep learning era. *arXiv preprint arXiv:2006.02419*, 2020. 1

[25] Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. Emergence of linguistic communication from referential games with symbolic and pixel input. In *International Conference on Learning Representations (ICLR)*, 2018. 1, 2, 4

[26] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-agent cooperation and the emergence of (natural) language. In *International Conference on Learning Representations (ICLR)*, 2017. 1, 2, 3, 4

[27] David K Lewis. *Convention: A Philosophical Study*. John Wiley & Sons, 1969. 5, 6

[28] Mengtian Li, Zhe Lin, Radomir Mech, Ersin Yumer, and Deva Ramanan. Photo-sketching: Inferring contour drawings from images. In *Proceedings of Winter Conference on Applications of Computer Vision (WACV)*, 2019. 3

[29] Daniela Mihai and Jonathon Hare. Learning to draw: Emergent communication through sketching. *arXiv preprint arXiv:2106.02067*, 2021. 3

[30] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013. 1, 3

[31] Igor Mordatch and Pieter Abbeel. Emergence of grounded compositional language in multi-agent populations. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, 2018. 1

[32] Umar Riaz Muhammad, Yongxin Yang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Learning deep sketch abstraction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[33] Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B. Cohen, and Simon Kirby. Compositional languages emerge in a neural iterated learning model. In *International Conference on Learning Representations (ICLR)*, 2020. 1, 2

[34] Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. The sketchy database: learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (TOG)*, 35(4):1–12, 2016. 2, 4, 5

[35] Bilge Sayim and Patrick Cavanagh. What line drawings reveal about the visual brain. *Frontiers in Human Neuroscience*, 5:118, 2011. 3

[36] David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In *International conference on machine learning*, pages 387–395. PMLR, 2014. 5

[37] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015. 4

[38] Jifei Song, Kaiyue Pang, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. Learning to sketch with shortcut cycle consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[39] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018. 2, 5, 11, 12

[40] Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pages 1057–1063, 2000. 5

[41] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11), 2008. 7

[42] Alexander Wang, Mengye Ren, and Richard Zemel. Sketchembednet: Learning novel concepts by imitating drawings. In *International Conference on Machine Learning*, pages 10870–10881. PMLR, 2021. 2

[43] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4):229–256, 1992. 6

[44] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015. 3

[45] Qian Yu, Feng Liu, Yi-Zhe Song, Tao Xiang, Timothy M Hospedales, and Chen-Change Loy. Sketch me that shoe. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[46] Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942, 2018. 6

[47] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 7

[48] Changqing Zou, Qian Yu, Ruofei Du, Haoran Mo, Yi-Zhe Song, Tao Xiang, Chengying Gao, Baoquan Chen, and Hao Zhang. Sketchyscene: Richly-annotated scene sketches. In *Proceedings of European Conference on Computer Vision (ECCV)*, 2018. 2

A. Category list

training categories				
apple	axe	bell	blimp	camel
cannon	car_(sedan)	chicken	cow	cup
deer	dolphin	duck	frog	giraffe
guitar	hamburger	horse	knife	mushroom
pig	pistol	pizza	rabbit	sailboat
seal	shark	sheep	snail	turtle
unseen categories				
pear	hammer	pickup truck	songbird	violin
sword	elephant	fish	penguin	swan

Table 2. Categories used in our game

We include 30 categories for training and 10 held-out categories for testing in our game; see Tab. 2.

B. Category embedding for other game settings

Fig. 7 shows the t-SNE visualization for other game settings. Agents under *max-step*, *sender-fixed*, and *one-step* settings fail to form clear boundaries between different categories, which makes it hard to observe semantic relations.

C. Visualizing sketch evolution

Visualizing the evolution process helps us understand what the agents have learned through communication regarding different categories. By comparing the evolved sketches with the intermediate results, we can know (i) how the agents abstract the sketch, (ii) which parts of the visual concept they highlight, and (iii) which parts are deemphasized. Figs. 8 to 10 show some evolution examples under different settings. Agents under *max-step* seem to abstract their drawings by repeatedly placing new strokes near old strokes, resulting in bold drawings. The number of strokes under *sender-fixed* gradually decreases, but the way of drawing will not change. Senders under *one-step* changes more wildly, but cannot form a consistent drawing behavior. Overall, compared with the *complete* setting, agents under the control settings do not form patterns to draw sketches, which echoes their relatively low classification results.

D. Learning objectives and training algorithm

Agents are trained jointly to maximize the objective:

$$\pi_S^*, \pi_R^* = \arg\max_{\pi_S, \pi_R} \mathbb{E}_{\tau\sim(\pi_S,\pi_R)}[\sum_{t=0} \gamma^t r_t], \tag{8}$$

where $\tau = \{C_0, a_{S0}, C_1, a_{R1}, a_{S1}, ...\}$ is the simulated episodic trajectory. To further expand the objective,

$$\begin{aligned}
&\mathbb{E}_{(\pi_S,\pi_R)}[\sum_{t=0} \gamma^t r_t] \\
=&\int p(I_S)p(I_R^1)...p(I_R^M)p(C_0) \\
&\cdot \int \pi_S(a_{S0}|I_S, C_0)\pi_R(a_{R1}|C_0, G(C_0, a_{S0}), I_R^1, ..., I_R^M) \\
&\cdot \left[r_0 + \mathbb{E}_{(\pi_S,\pi_R)}\left[\sum_{t=1} \gamma^t r_t\right]\right] da_{S0} da_{R1} dI_S dI_R^1 ... dI_R^M dC_0 \\
=&\mathbb{E}_{I_S, I_R^1, ..., I_R^M, C_0}\left[\mathbb{E}_{(\pi_S,\pi_R)}\left[r_0 + \mathbb{E}_{(\pi_S,\pi_R)}\left[\sum_{t=1} \gamma^t r_t\right]\right]\right]
\end{aligned} \tag{9}$$

We calculate $\mathbb{E}_{I_S, I_R^1, ..., I_R^M, C_0}[\cdot]$ by sampling I_S, I_R, and initializing C_0 to blank at each round. We represent the $\mathbb{E}_{(\pi_S,\pi_R)}[\cdot]$ as $\mathcal{V}(X_0)$ and use $V_\lambda(X_1)$ to estimate the reward expectation $\mathbb{E}_{(\pi_S,\pi_R)}[\sum_{t=1} \gamma^t r_t]$:

$$\begin{aligned}
\mathcal{V}(X_0) = &\mathbb{E}_{(\pi_S(a_{S0}|I_S,C_0), \pi_R(a_{R1}|C_0, G(C_0,a_{S0}), I_R^1, ..., I_R^M))} \\
&\cdot[(r_0 + \gamma\delta(a_{R1})V_\lambda(X_1)],
\end{aligned} \tag{10}$$

where $X_t = [I_S, I_R^1, ..., I_R^M, C_t, C_{t+1}], t = 0, 1..., \ \delta(\cdot)$ is the Dirac delta function that returns 1 when the action is *wait* and 0 otherwise.

The sender policy is parametrized as a Gaussain distribution,

$$\pi_S = \mathcal{N}(\mu_t, \sigma^2), \quad \mu_t = h_S(I_S, C_t), \quad \sigma^2 = c \cdot \mathbf{I}, \tag{11}$$

such that a_{S0} can be written as

$$a_{S0} = \mu_0 + \sigma\epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I}). \tag{12}$$

Therefore, we can expand $\mathcal{V}(X_0)$ as,

$$\begin{aligned}
\mathcal{V}(X_0) = &\int \pi_S(a_{S0}|C_0, I_S)\mathbb{E}_{\pi_R(a_{R1}|C_0, G(C_0,a_{S0}), I_R^1, ..., I_R^M)} \\
&\cdot[r_0 + \gamma\delta(a_{R1})V_\lambda(X_1)]da_{S0} \\
= &\int p(\epsilon)\mathbb{E}_{\pi_R(a_{R1}|C_0, G(C_0, \mu_0+\sigma\epsilon), I_R^1, ..., I_R^M)} \\
&\cdot[r_0 + \gamma\delta(a_{R1})V_\lambda(X_1)]d\epsilon \\
= &\mathbb{E}_\epsilon[\mathbb{E}_{\pi_R}[r_0 + \gamma\delta(a_{R1})V_\lambda(X_1)]]
\end{aligned} \tag{13}$$

$\mathbb{E}_\epsilon[\cdot]$ is approximated with a point estimate. Since π_R is a categorical distribution, we expand $\mathbb{E}_{\pi_R}$ as

$$\begin{aligned}
&\mathbb{E}_{\pi_R}[r_0 + \gamma\delta(a_{R1})V_\lambda(X_1)] \\
= &\sum_{j=1}^{M+1} p(a_{R1}^j)[r_0^j + \gamma\delta(a_{R1})V_\lambda(X_1)].
\end{aligned} \tag{14}$$

$V_\lambda(X_t)$ in Eq. (5) is an eligibility trace approximation of the ground-truth value function [39]. Considering the early

Figure 7. **t-SNE of visual embedding**. These embeddings are extracted from the finetuned VGGNet used for evolved sketch classification under the *max-step* (left), *sender-fixed* (middle), and *one-step* (right) setting, respectively. Neither of them forms a clear boundary between different categories.

termination in our setting, we set the time step when the receiver makes the prediction as T_{choice}. When t is the time step less or equal than T_{choice}, V_λ mixes Monte Carlo estimate at different roll-out lengths. Otherwise, we only have an estimated value $v_\phi(X_t)$.

$$V_\lambda(X_t) = \begin{cases} (1-\lambda)\sum_{n=1}^{H-1}\lambda^{n-1}V_N^n(X_t) + \lambda^{H-1}V_N^H(X_t) \\ \qquad\qquad\qquad\qquad \text{if } t \leqslant T_{\text{choice}} \\ v_\phi(X_t) \qquad\qquad\qquad \text{otherwise} \end{cases} \tag{15}$$

where $H = T_{\text{choice}} - t + 1$, and $V_N^k(X_t)$ is the Monte Carlo estimate at k roll-out lenghts. $V_N^k(X_t) = \mathbb{E}_{\pi_S,\pi_R}[\sum_{n=t}^{h-1}\gamma^{n-t}r_n + \gamma^{h-t}\delta(a_{Rh})v_\phi(X_h)]$, with $h = \min(t+k, T_{\text{choice}})$ being the maximal timestep. Due to the error reduction property [39], the eligibility trace estimation $V_\lambda(\cdot)$ is less biased than $v_\phi(\cdot)$. When regressing $v_\phi(X_t)$ towards the bootstrapped $V_\lambda(X_t)$,

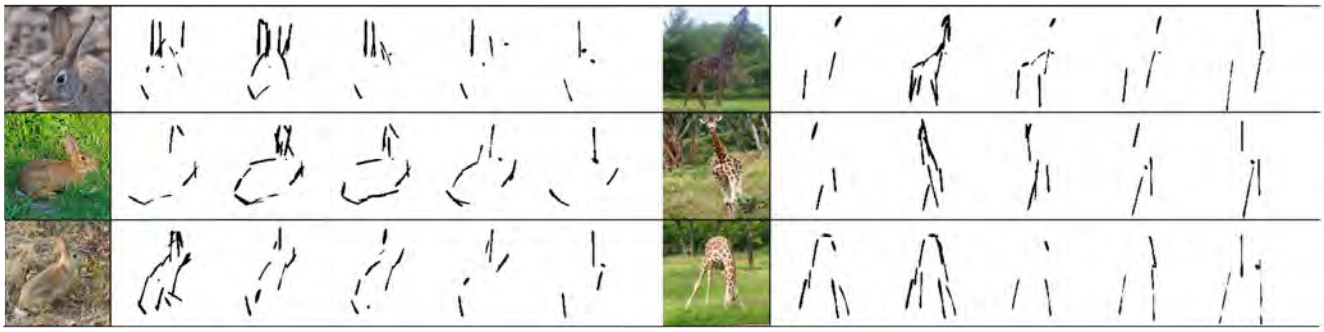
$$\phi^* = \arg\max_\phi \mathbb{E}_{\pi_S,\pi_R}[\sum\nolimits_t \frac{1}{2}||v_\phi(X_t) - V_\lambda(X_t)||^2]. \tag{16}$$

$v_\phi(X_t)$ will be improved towards the fixed point. The training algorithm is shown in Algorithm 1.

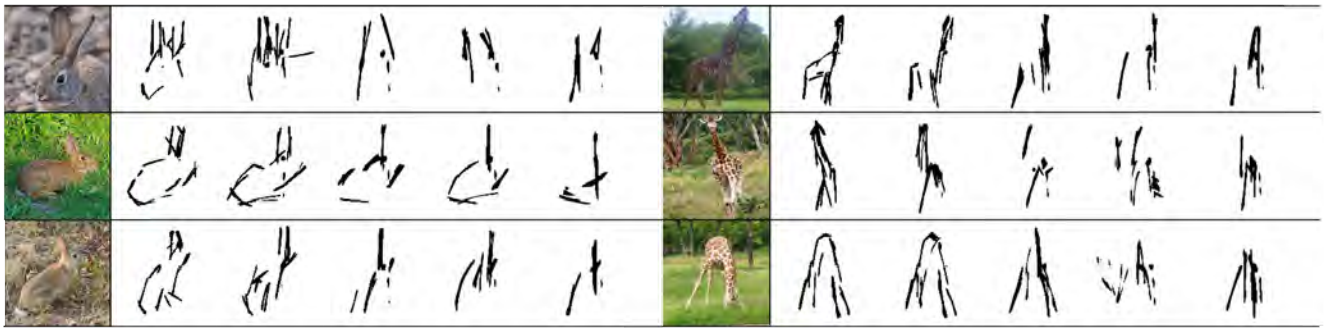
Algorithm 1: Training Algorithm

Initialization: Initialize neural network parameters θ, ρ, ϕ for π_S, π_R and v_ϕ respectively

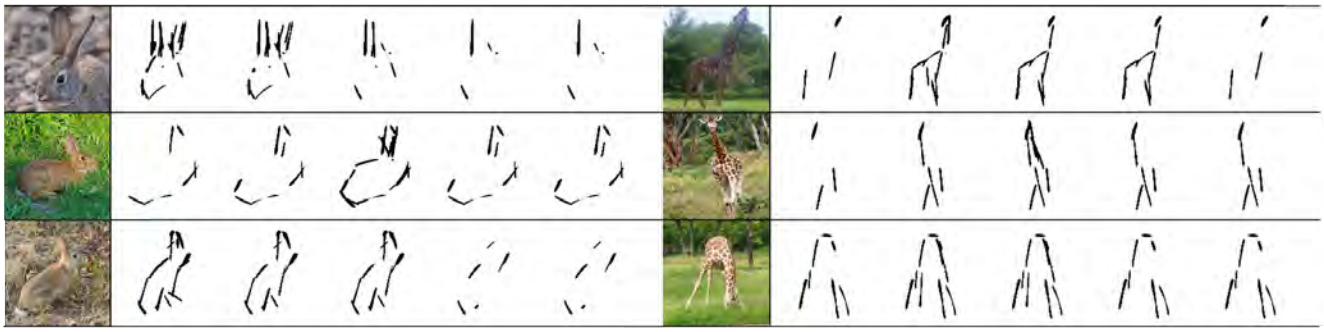
for *game round* $l = 1, ..., L$ **do**
  for *time step* $t = 0, ..., T$ **do**
    $a_{St} \sim \pi_S(a_{St}|C_t, I_S)$
    $C_{t+1} = G(C_t, a_{St})$
    $a_{Rt+1} \sim \pi_R(a_{Rt+1}|C_t, C_{t+1}, I_R^1, ..., I_R^M)$
    if a_{Rt+1} *is not wait* **then**
      $T_{\text{choice}} = t$
    end
  end
  Compute $\{V_\lambda(X_t)\}_{t=1}^T$ via Eq. (15)
  Compute $\{\mathcal{V}(X_t)\}_{t=1}^T$ via Eq. (10)
  Update $\theta \leftarrow \theta + \alpha_S \nabla_\theta \sum_t \mathcal{V}(X_t)$
  Update $\rho \leftarrow \rho + \alpha_R \nabla_\rho \sum_t \mathcal{V}(X_t)$
  Update $\phi \leftarrow \phi - \alpha_v \nabla_\phi \sum_t \frac{1}{2}||v_\phi(X_t) - V_\lambda(X_t)||^2$
end



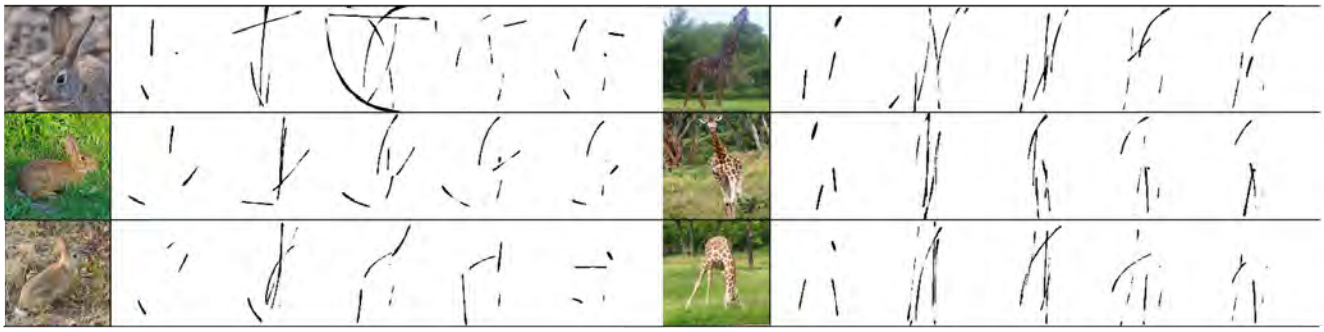
(a) *complete* example 1



(b) *max-step* example 1

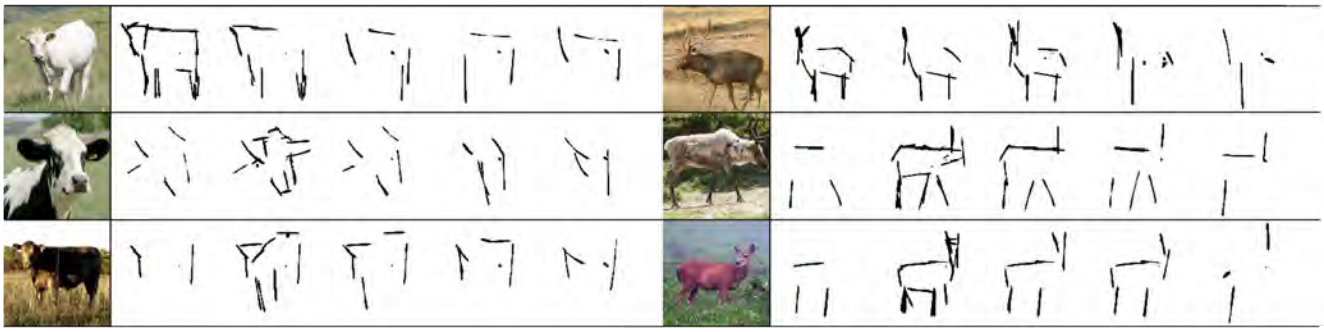


(c) *sender-fixed* example 1

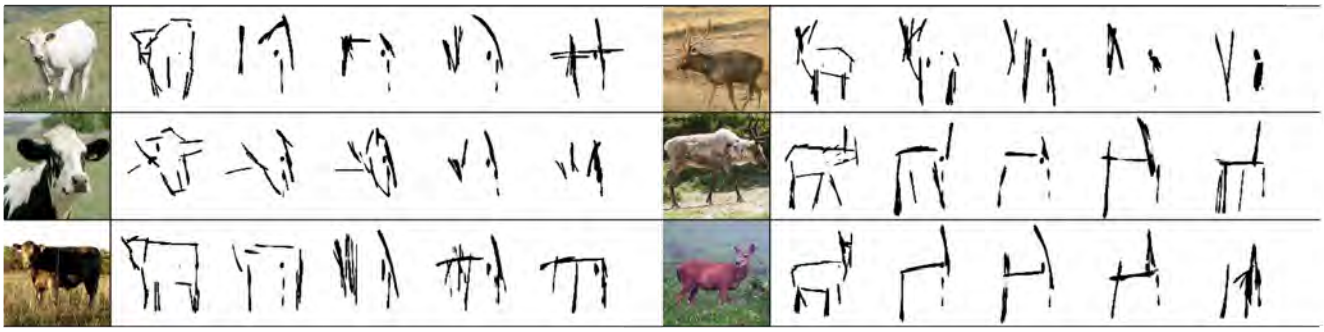


(d) *one-step* example 1

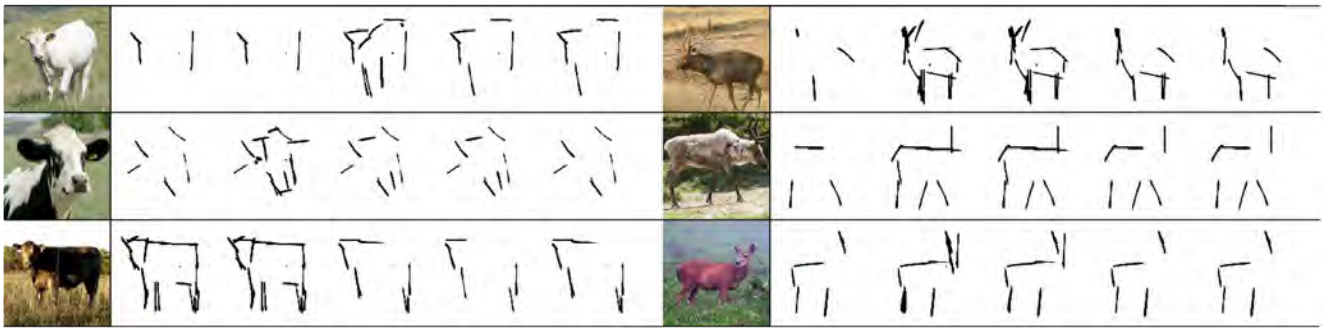
Figure 8. **Evolution of *rabbit* and *giraffe* under different settings.** Compared to other settings, agents under *complete* setting consistently highlight the ears of *rabbit* and the neck of *giraffe*.



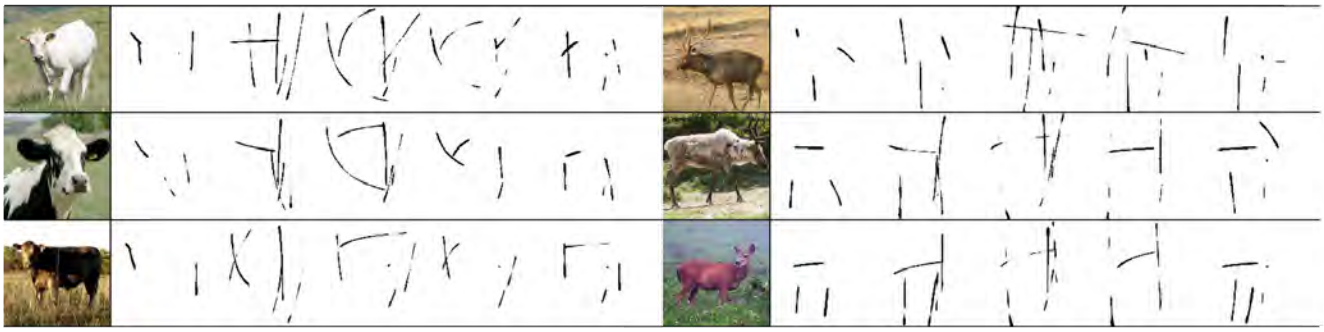
(a) *complete* example 2



(b) *max-step* example 2

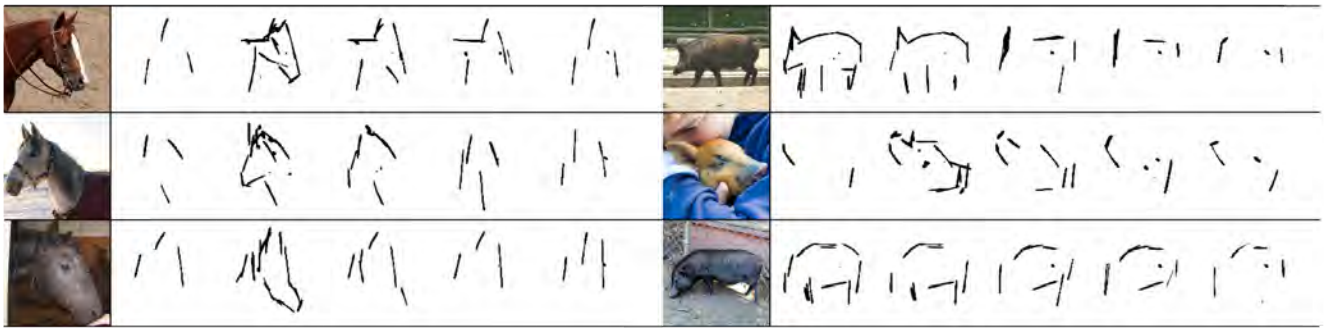


(c) *sender-fixed* example 2

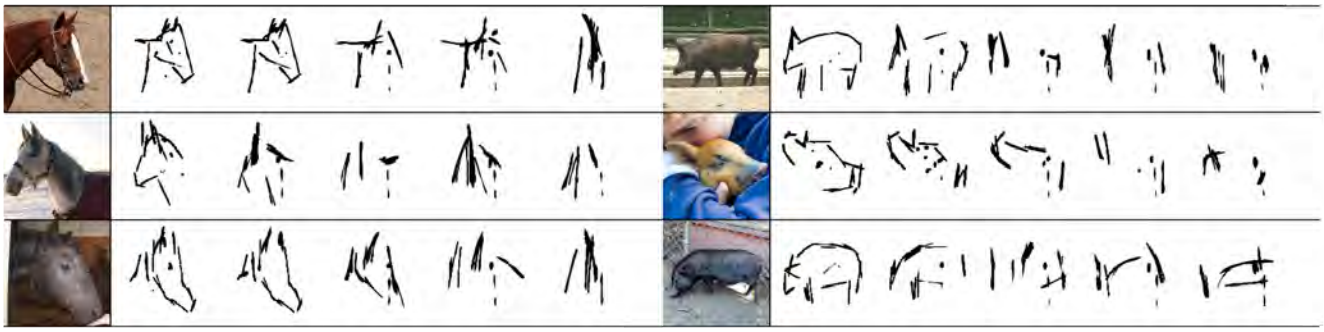


(d) *one-step* example 2

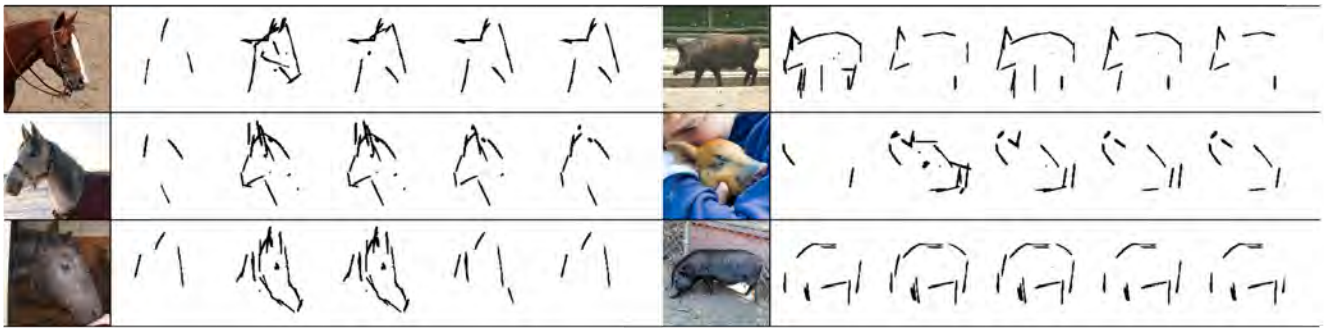
Figure 9. **Evolution of *cow* and *deer* under different settings.** The sketches of *cow* all form a "horn" shape at the left under *complete* setting, whereas others do not form this pattern. In *complete* setting, the sketches of *deer* converge to emphasize the antler of the deer. Some sketches under other settings also show a vertical line, but ones in the *complete* are more consistent.



(a) *complete* example 3



(b) *max-step* example 3



(c) *sender-fixed* example 3

(d) *one-step* example 3

Figure 10. **Evolution of *horse* and *pig* under different settings.** In the *complete* setting, sketches of *horse* all show three vertical lines. For different instances of *pig*, agents all draw a single line on the right. We do not obverse obvious patterns in other settings.