# *Supplementary Materials* for Emergent Graphical Conventions in a Visual Communication Game

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

[1] Department of Computer Science, UCLA
[2] Beijing Institute for General Artificial Intelligence (BIGAI)
[3] Department of Communication, UCLA [4] Department of Statistics, UCLA
[5] Institute for Artificial Intelligence, Peking University

https://sites.google.com/view/emergent-graphical-conventions

## A Category list

Table A1: **Categories used in our game**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 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 | | | | | | |

We include 30 categories for training and 10 held-out categories for testing in our game; see Tab. A1.

## B Category embedding for other game settings

Fig. A1 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 Learning objectives and training algorithm

Agents are trained jointly to maximize the objective:

$$\pi_S^*, \pi_R^* = \arg\max_{\pi_S,\pi_R} \mathbb{E}_{\tau\sim(\pi_S,\pi_R)}[\sum_{t=0} \gamma^t r_t], \tag{A1}$$

⋆ indicates equal contribution.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

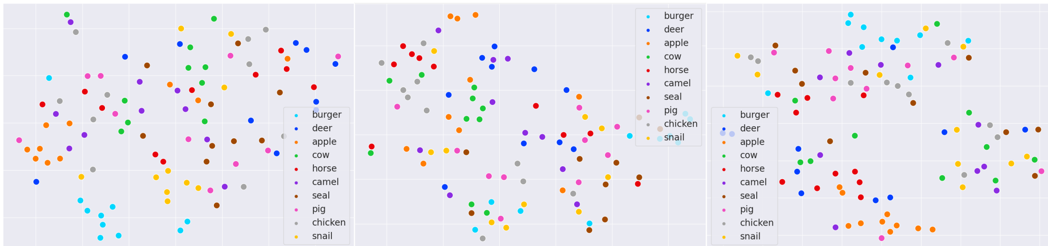


Figure A1: **t-SNE of visual embedding**. These embeddings are extracted from the finetuned VGGNet used for evolved sketch classification under the *max-step* (left), *sender-fixed* (middle), and *one-step* (right) settings, respectively. Neither of them forms a clear boundary between different categories.

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

$$\begin{aligned}\mathbb{E}_{(\pi_S,\pi_R)}[\sum_{t=0}\gamma^t r_t] = &\int p(I_S)p(I_R^1)...p(I_R^M)p(C_0)\\ &\int \pi_S(a_{S0}|I_S, C_0)\pi_R(a_{R1}|C_0, G(C_0, a_{S0}), I_R^1, ..., I_R^M)\\ &\cdot \left[r_0 + \mathbb{E}_{(\pi_S,\pi_R)}\left[\sum_{t=1}\gamma^t r_t\right]\right] da_{S0}da_{R1}dI_S dI_R^1...dI_R^M dC_0\\ &= \mathbb{E}_{I_S, I_R^1, ..., I_R^M, C_0}\left[\mathbb{E}_{(\pi_S,\pi_R)}\left[r_0 + \mathbb{E}_{(\pi_S,\pi_R)}\left[\sum_{t=1}\gamma^t r_t\right]\right]\right]\end{aligned} \tag{A2}$$

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 $\mathbb{E}_{(\pi_S,\pi_R)}[\sum_{t=1}\gamma^t r_t]$:

$$\mathcal{V}(X_0) \quad = \quad \mathbb{E}_{(\pi_S(a_{S0}|I_S,C_0),\pi_R(a_{R1}|C_0,G(C_0,a_{S0}),I_R^1,...,I_R^M))}[(r_0 \quad + \quad \gamma\delta(a_{R1})V_\lambda(X_1)], \tag{A3}$$

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 Gaussian distribution,

$$\pi_S = \mathcal{N}(\mu_t, \sigma^2), \quad \mu_t = h_S(I_S, C_t), \quad \sigma^2 = c \cdot \mathbf{I}, \tag{A4}$$

such that $a_{S0}$ can be written as

$$a_{S0} = \mu_0 + \sigma\epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I}). \tag{A5}$$

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

$$\begin{aligned}\mathcal{V}(X_0) &= \int \pi_S(a_{S0}|C_0, I_S)\mathbb{E}_{\pi_R(a_{R1}|C_0,G(C_0,a_{S0}),I_R^1,...,I_R^M)} \cdot [r_0 + \gamma\delta(a_{R1})V_\lambda(X_1)]da_{S0}\\ &= \int p(\epsilon)\mathbb{E}_{\pi_R(a_{R1}|C_0,G(C_0,\mu_0+\sigma\epsilon),I_R^1,...,I_R^M)} \cdot [r_0 + \gamma\delta(a_{R1})V_\lambda(X_1)]d\epsilon\\ &= \mathbb{E}_\epsilon[\mathbb{E}_{\pi_R}[r_0 + \gamma\delta(a_{R1})V_\lambda(X_1)]]\end{aligned} \tag{A6}$$

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

$$\mathbb{E}_{\pi_R}[r_0 + \gamma\delta(a_{R1})V_\lambda(X_1) = \sum_{j=1}^{M+1} p(a_{R1}^j)[r_0^j + \gamma\delta(a_{R1})V_\lambda(X_1)]. \tag{A7}$$

$V_\lambda(X_t)$ in Eq. (A3) is an eligibility trace approximation of the ground-truth value function (Sutton and Barto, 2018). Considering the early termination in our setting, we set the time step when the receiver makes the prediction as $T_{\text{choice}}$. When $t$ is the time step less or equal than $T_{\text{choice}}$, $V_\lambda$ mixes

Monte Carlo estimate at different roll-out lengths. Otherwise, we only have an estimated value $v_\phi(X_t)$.

$$V_\lambda(X_t) = \begin{cases} (1-\lambda)\sum_{n=1}^{H-1}\lambda^{n-1}V_N^n(X_t) + \lambda^{H-1}V_N^H(X_t) & \\ \qquad\qquad\qquad\qquad\qquad \text{if } t \le T_{\text{choice}} \\ v_\phi(X_t) \qquad\qquad\qquad\qquad \text{otherwise} \end{cases} \tag{A8}$$

where $H = T_{\text{choice}} - t + 1$, and $V_N^k(X_t)$ is the Monte Carlo estimate at $k$ roll-out lengths. $V_N^k(X_t) = \mathbb{E}_{\pi_S,\pi_R}[\sum_{n=t}^{h-1}\gamma^{n-t}r_n + \gamma^{h-t}\delta(a_{Rh})v_\phi(X_h)]$, with $h = \min(t+k, T_{\text{choice}})$ being the maximal timestep. Due to the error reduction property Sutton and Barto (2018), the eligibility trace estimation $V_\lambda(\cdot)$ is less biased than $v_\phi(\cdot)$. When regressing $v_\phi(X_t)$ towards the bootstrapped $V_\lambda(X_t)$,

$$\phi^* = \arg\max_\phi \mathbb{E}_{\pi_S,\pi_R}[\sum\nolimits_t \frac{1}{2}||v_\phi(X_t) - V_\lambda(X_t)||^2]. \tag{A9}$$

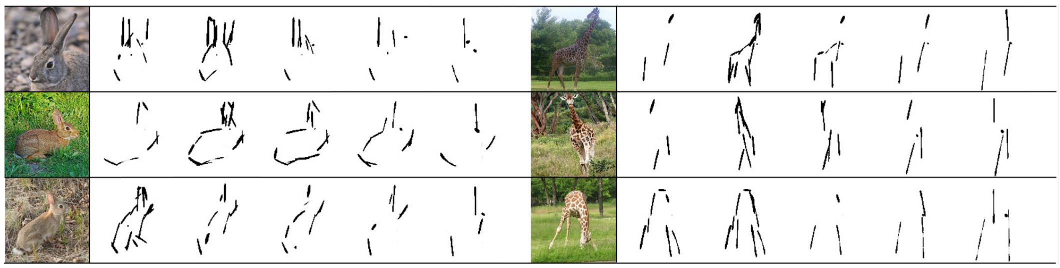
$v_\phi(X_t)$ will be improved towards the fixed point.

## D Visualizing sketch evolution

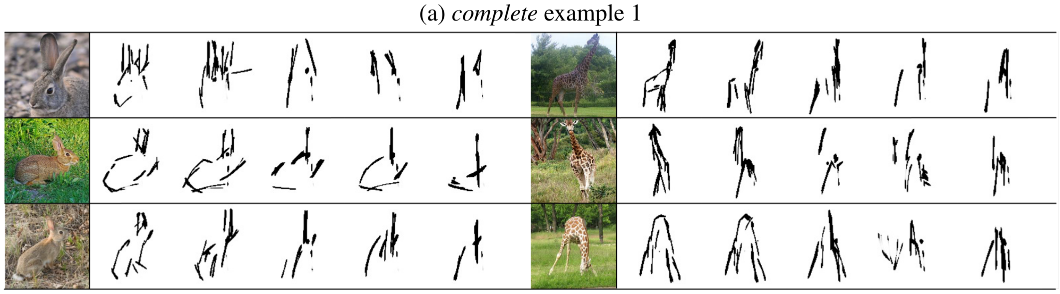
Visualizing the evolution process helps us understand what the agents have learned through communication regarding different categories. By comparing the evolved sketches with the intermediate results, we can know (i) how the agents abstract the sketch, (ii) which parts of the visual concept they highlight, and (iii) which parts are de-emphasized. Fig. A2 to A4 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 the drawing will not change. Senders under *one-step* change 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.

## References

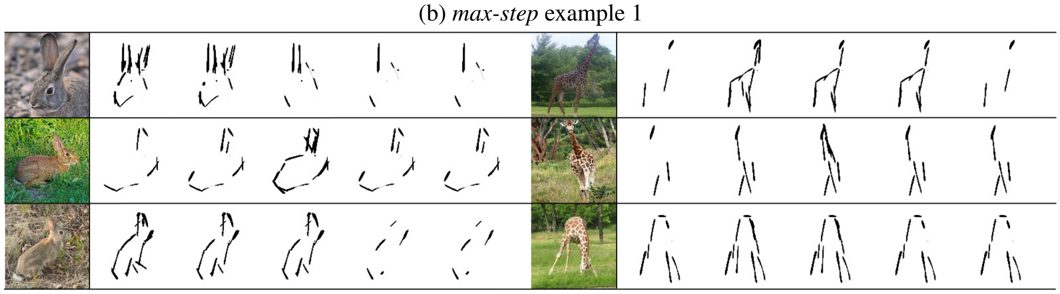
Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press. 2, 3



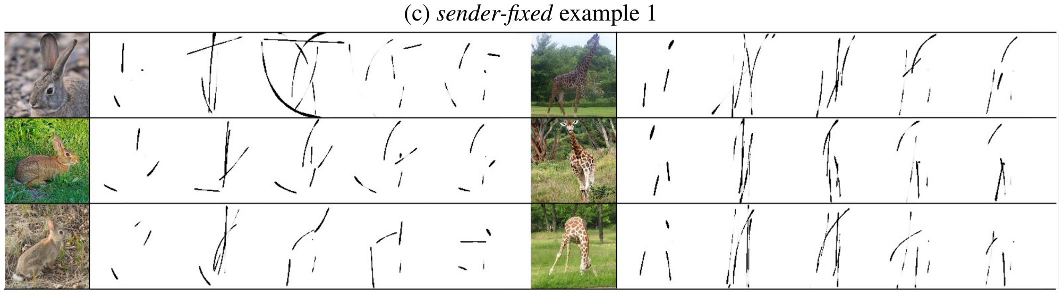
(a) *complete* example 1



(b) *max-step* example 1

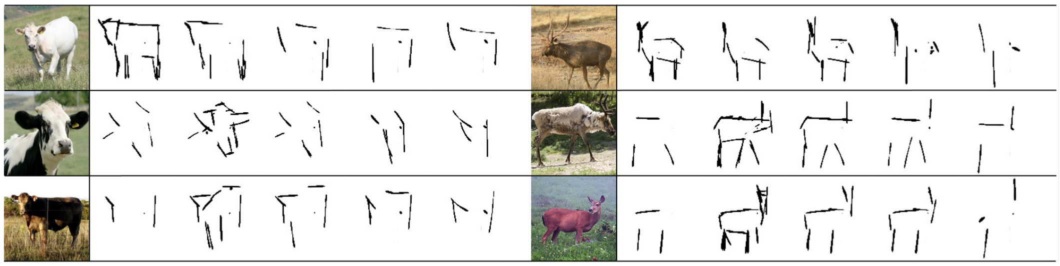


(c) *sender-fixed* example 1

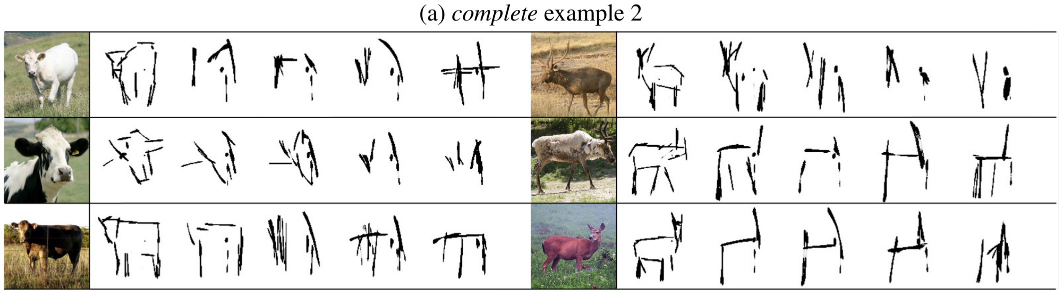


(d) *one-step* example 1

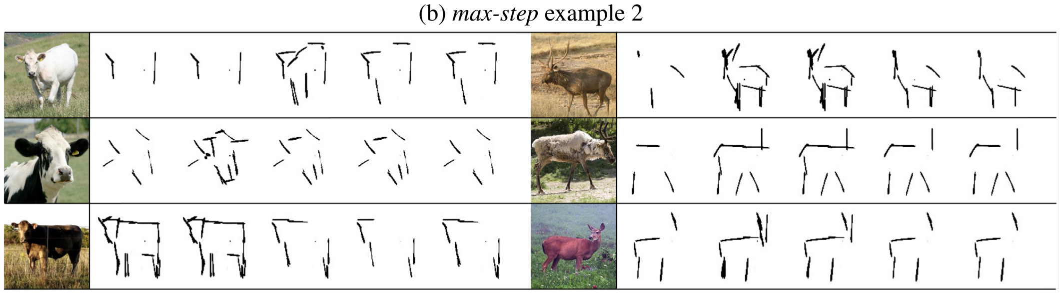
Figure A2: **Evolution of *rabbit* and *giraffe* under different settings.** Compared to other settings, agents under *complete* setting consistently highlight the ears of *rabbit* and the neck of *giraffe*.



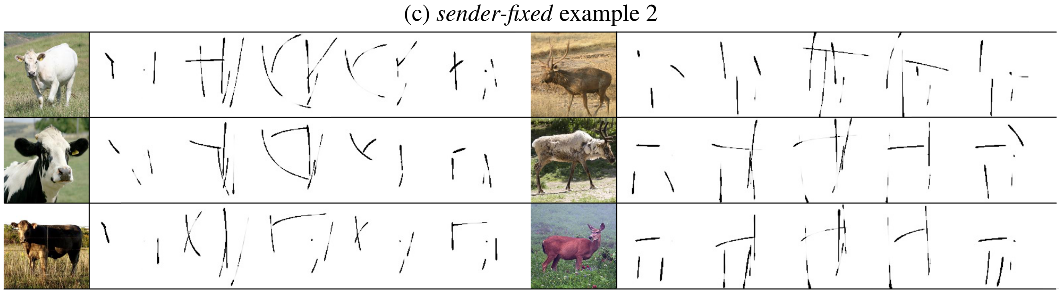
(a) *complete* example 2



(b) *max-step* example 2

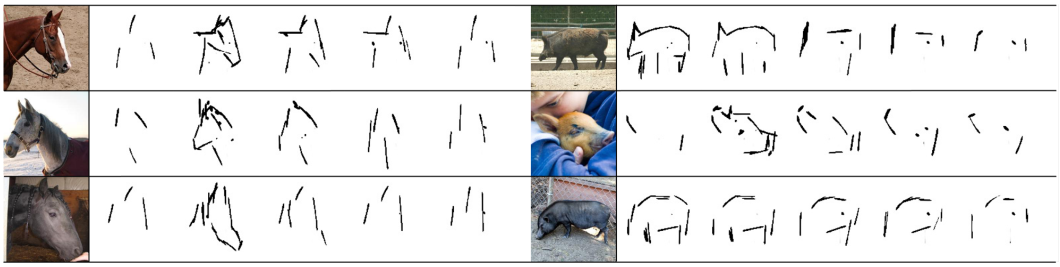


(c) *sender-fixed* example 2

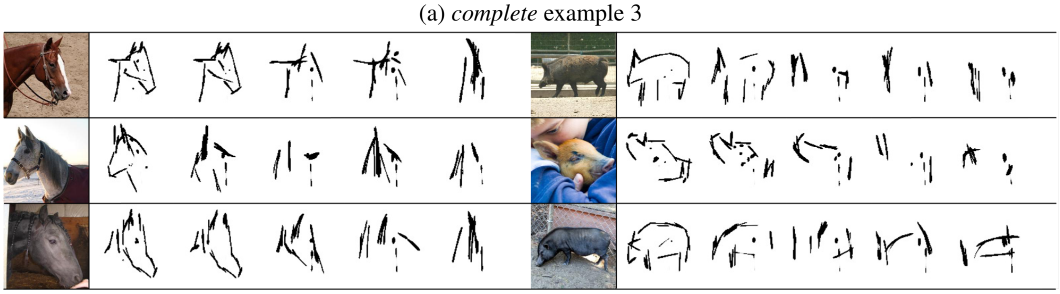


(d) *one-step* example 2

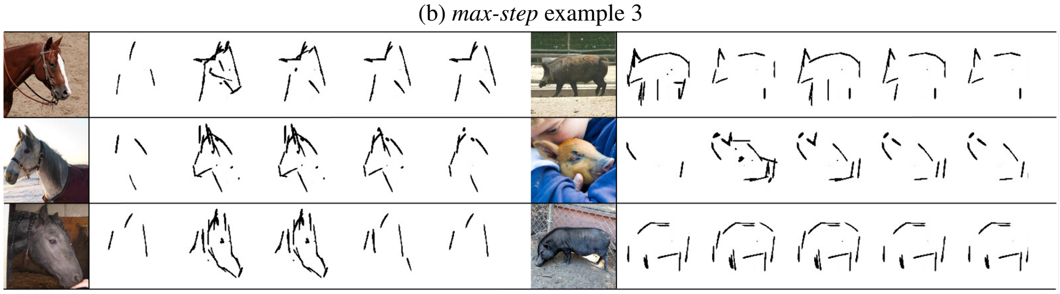
Figure A3: **Evolution of *cow* and *deer* under different settings.** The sketches of *cow* all form a "horn" shape at the left under *complete* setting, whereas others do not form this pattern. In *complete* setting, the sketches of *deer* converge to emphasize the antler of the deer. Some sketches under other settings also show a vertical line, but the ones in the *complete* are more consistent.



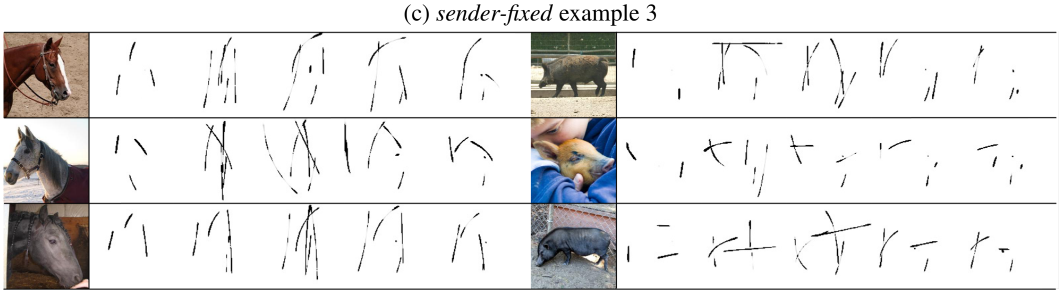
(a) *complete* example 3



(b) *max-step* example 3



(c) *sender-fixed* example 3



(d) *one-step* example 3

Figure A4: **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.