A multi-modal, multi-agent system

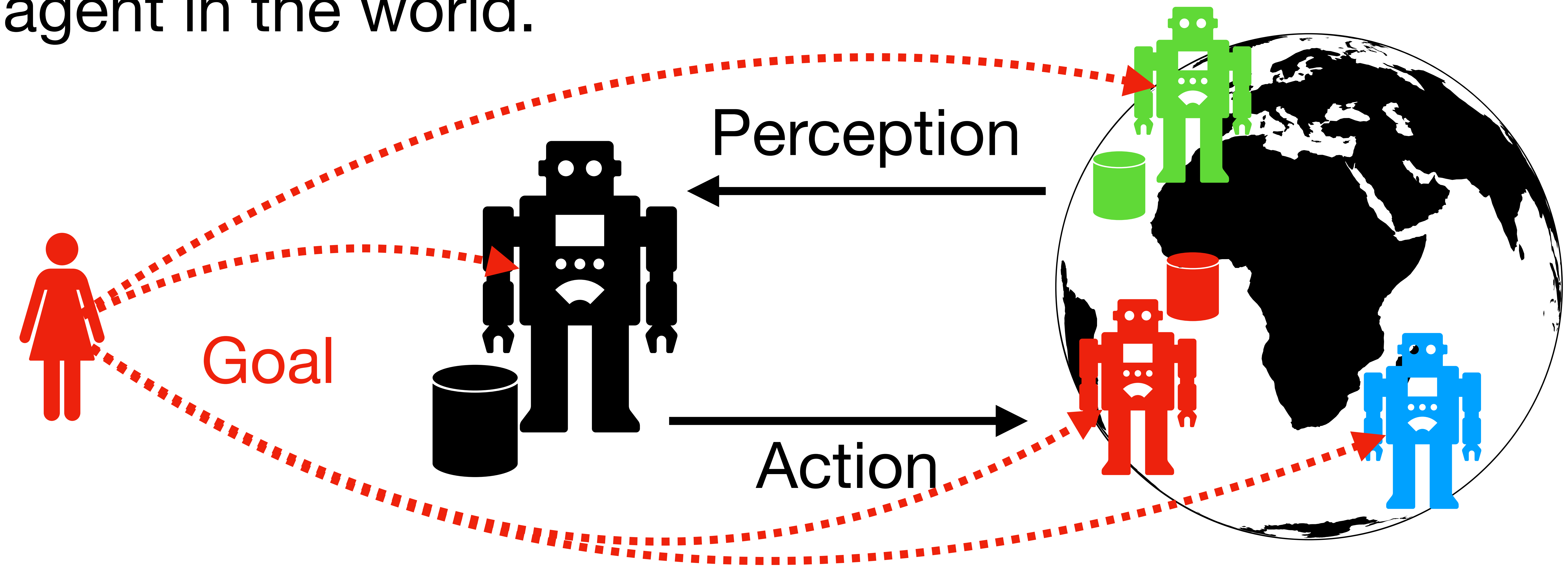
A minimalistic framework to study multi-agent systems

Kyunghyun Cho with Katrina Drozdov, Andrew Drozdov and Douwe Kiel

On a job market!

A minimalistic framework to study multi-agent systems

- The world can be perceived in different ways.
- An agent perceives parts of and acts on the world.
- An agent works toward a goal.
- There are more than one agent in the world.
- Agents are symmetric.



A minimalistic framework to study multi-agent systems

- The world can be perceived in different ways.

llama noun

lla·ma ˈlä-mə ˈyä-mə

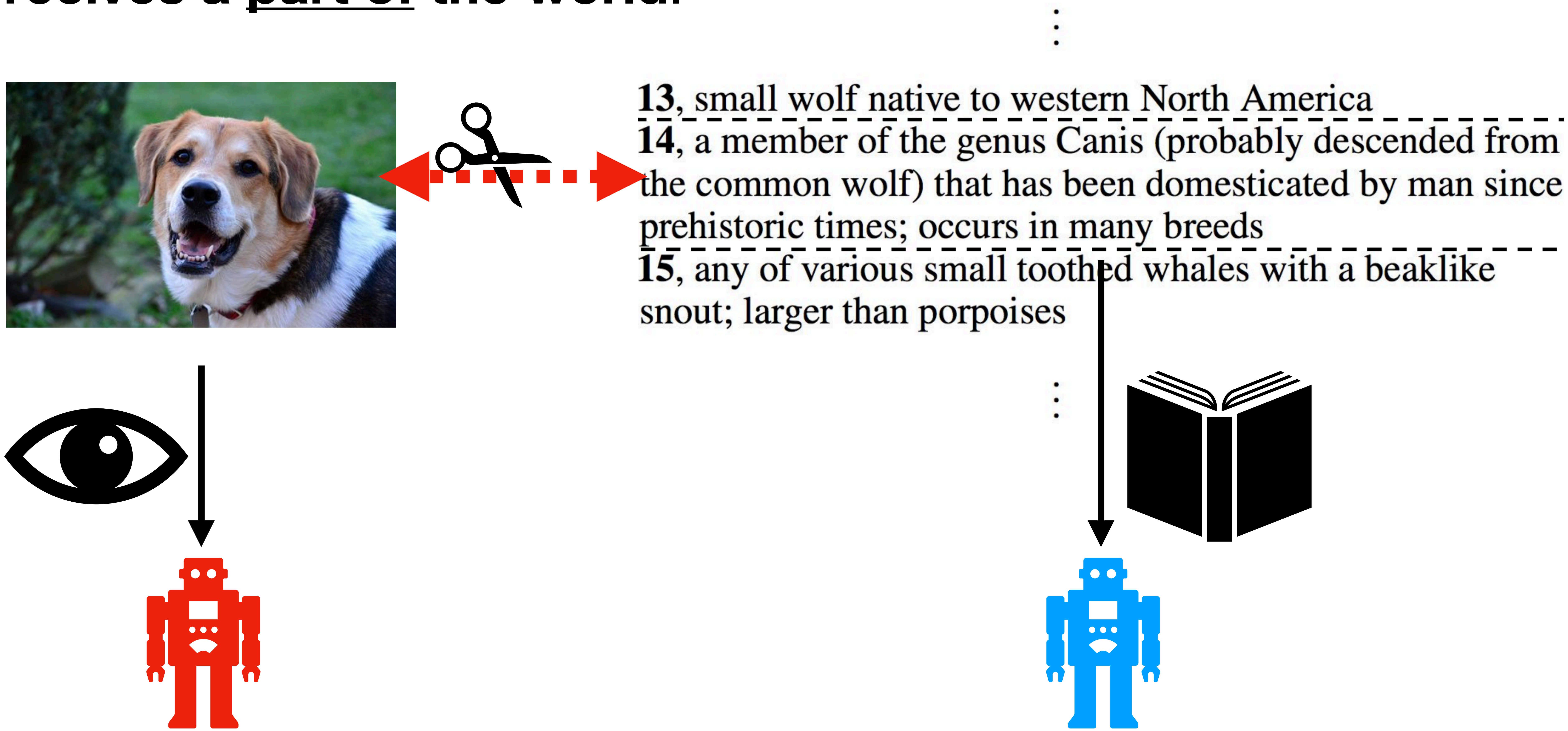
: any of a genus (*Lama*) of wild or domesticated, long-necked, South American ruminant (see RUMINANT entry 1) mammals related to the camels but smaller and without a hump

especially **:** a domesticated llama (*L. glama*) descended from the guanaco and used especially in the Andes as a pack animal and a source of wool

Merriam-Webster
Est. 1828

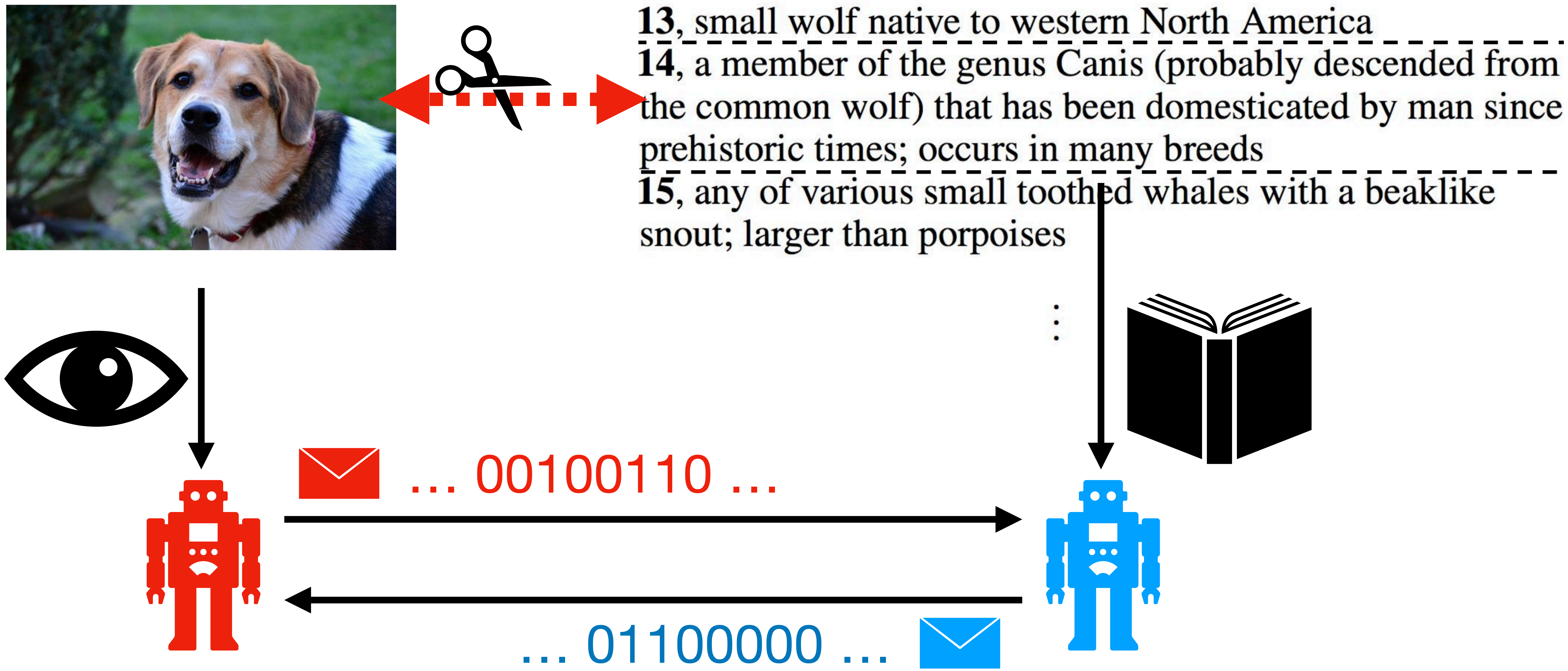
A minimalistic framework to study multi-agent systems

- An agent **perceives a <u>part of</u> the world**.



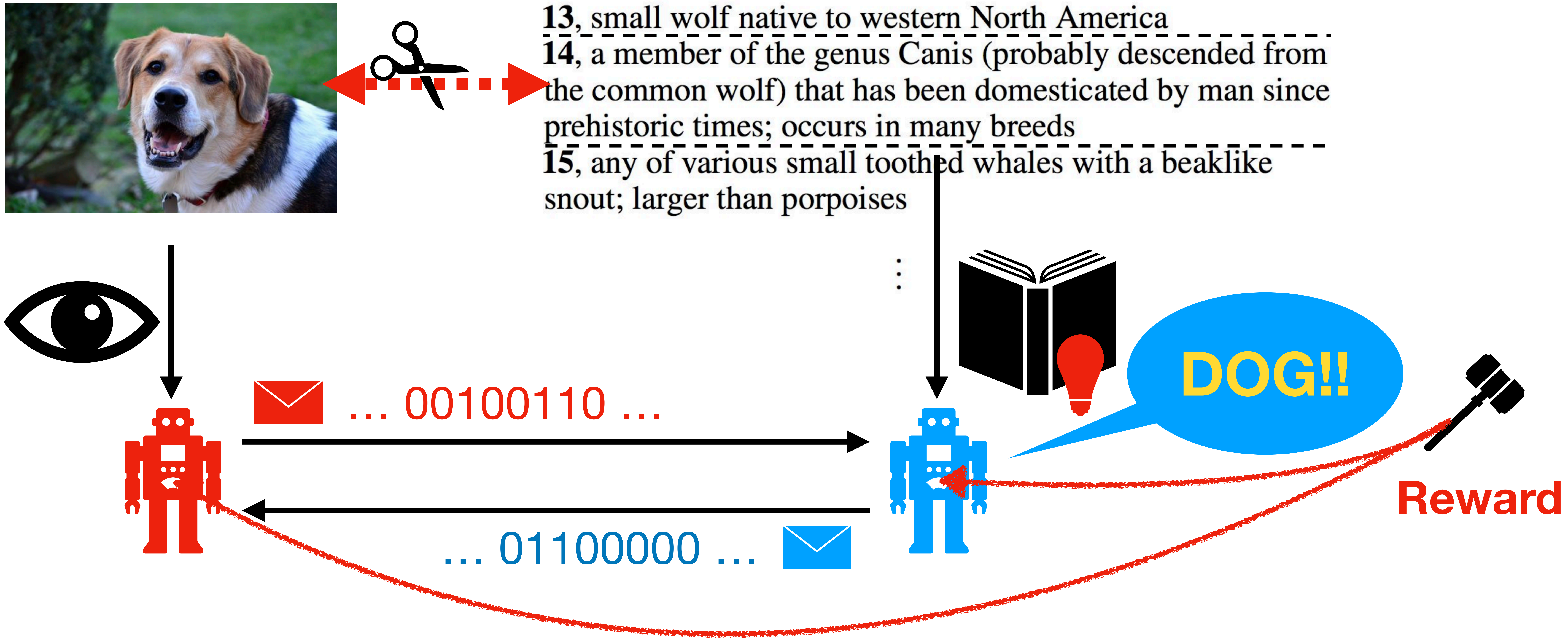
A minimalistic framework to study multi-agent systems

- An agent **acts on the world**: In this case, by communicating with each other.



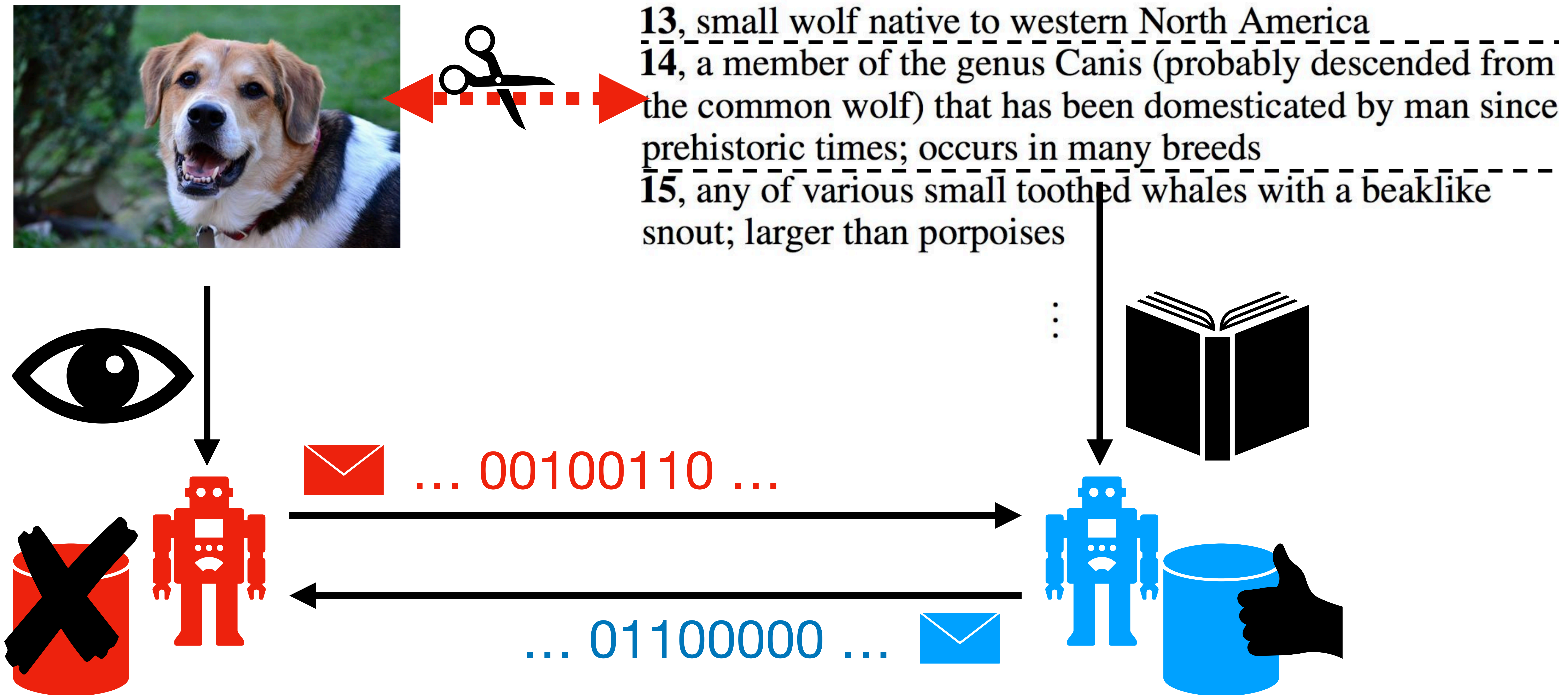
A minimalistic framework to study multi-agent systems

- An agent works toward a **shared goal**: figure out the missing connection.



A minimalistic framework to study multi-agent systems

- Agents are **asymmetric**: the visual agent has no memory, but the reader has one.

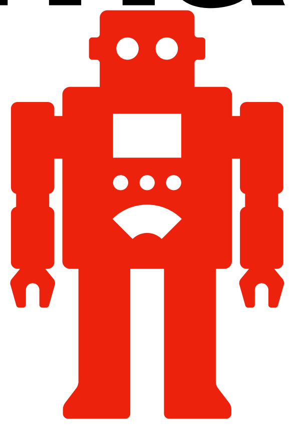


A minimalistic framework to study multi-agent systems

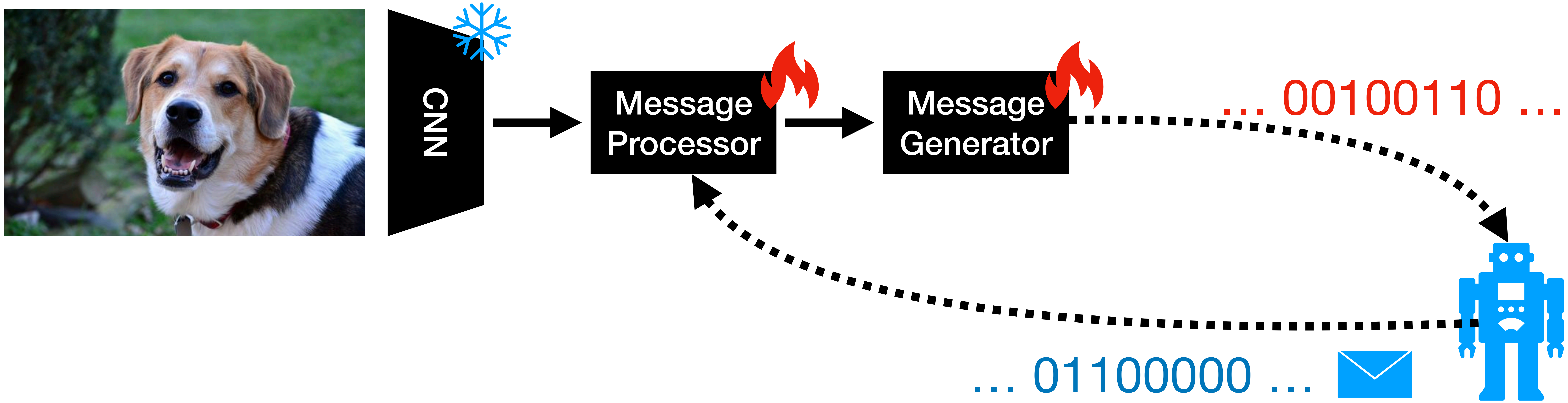
- This is often called a referential/signaling game [Lewis et al., 1969].
- This minimal framework allows us to test:
    - Multi-modal processing capabilities: can agents handle images and text?
    - Test-time compute: can the conversation grows to solve challenging cases?
    - Reinforcement learning: can the agents develop a chain of discrete messages end-to-end to maximize a joint reward?

A minimalistic implementation

A Sender

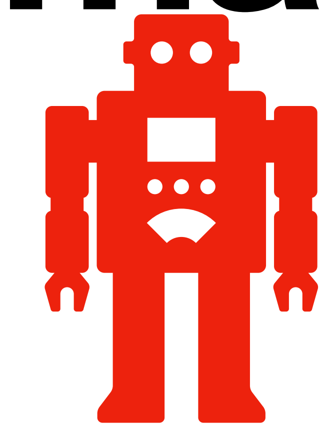


- The sender does not have a memory.
- The sender uses a pretrained convolutional network for image processing.
- It learns to process the visual information and the received message.
- It then produces a bit string as a message to be sent to the receiver.



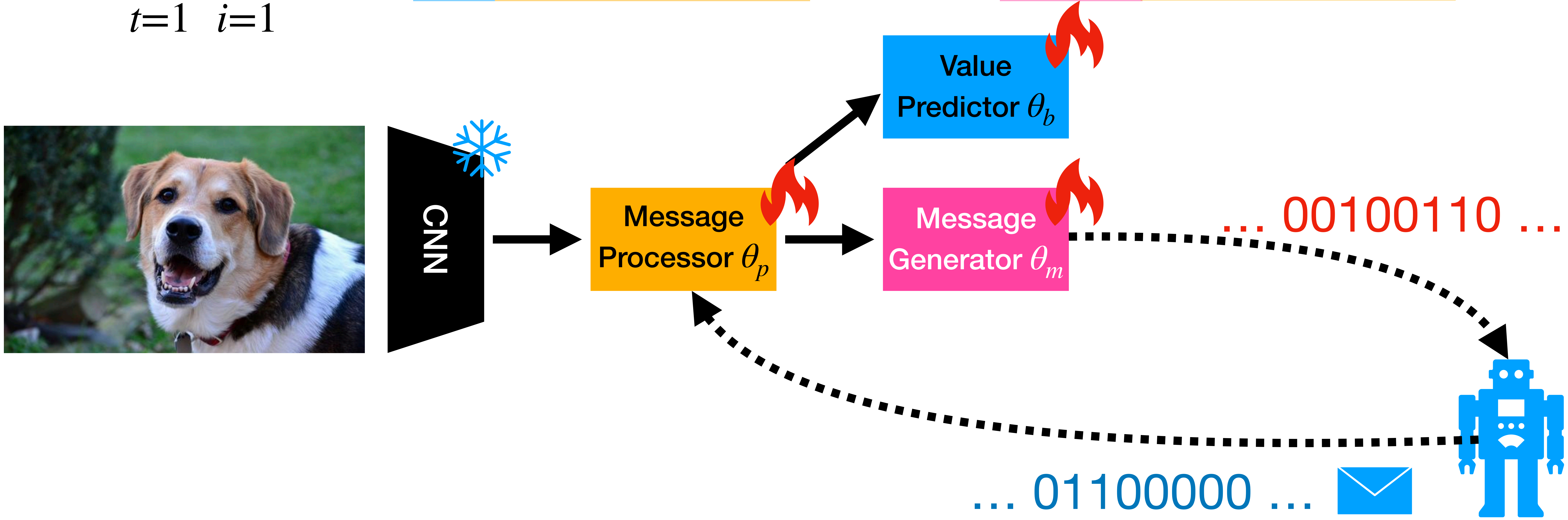
A minimalistic implementation

A Sender



- We use vanilla policy gradient (REINFORCE) with a baseline technique.

$$\nabla_{\theta_m} = \sum_{t=1}^{T} \sum_{i=1}^{L} (R - B_s(F(X_{\text{img}}, m_r^{t-1}))) \nabla_m \log p(m_i^t \mid F(X_{\text{img}}, m_r^{t-1})))$$

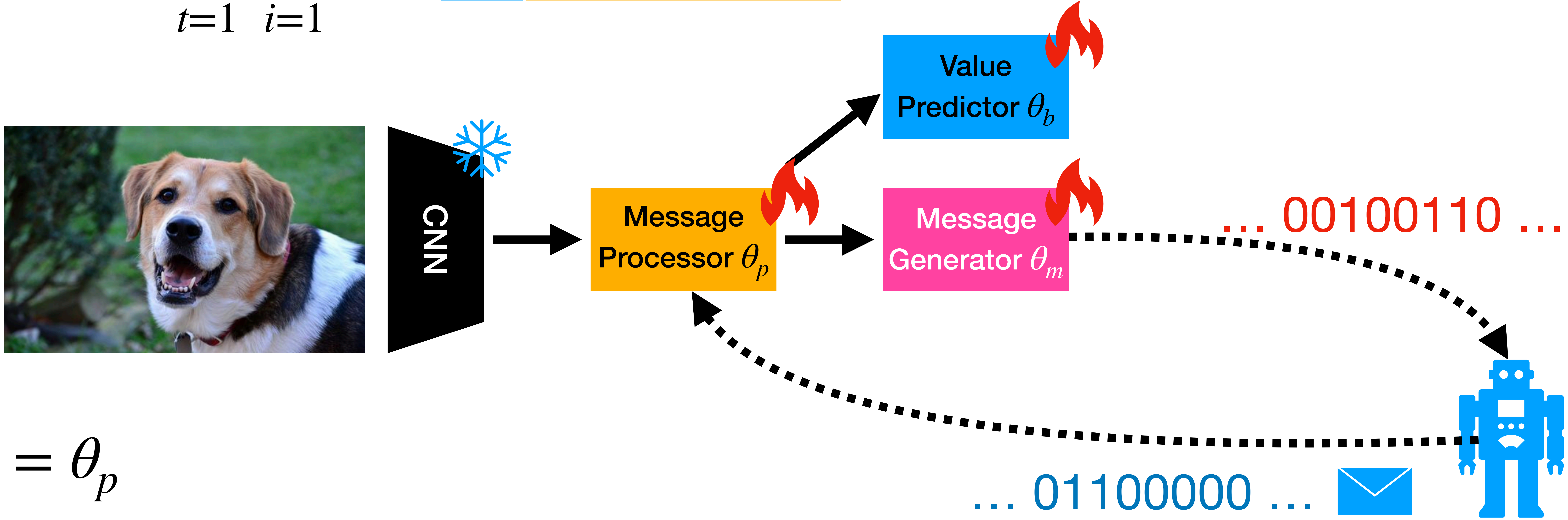


A minimalistic implementation

A Sender

- The baseline (or value) predictor is trained to predict the overall reward.

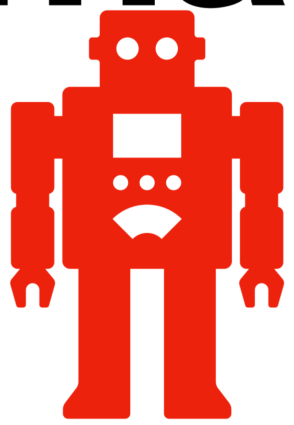
$$\nabla_{\theta_b} = -\sum_{t=1}^{T}\sum_{i=1}^{L}\left(R - B_s(F(X_{\text{img}}, m_r^{t-1}))\right)\nabla_b B_s(F(X_{\text{img}}, m_r^{t-1}))$$



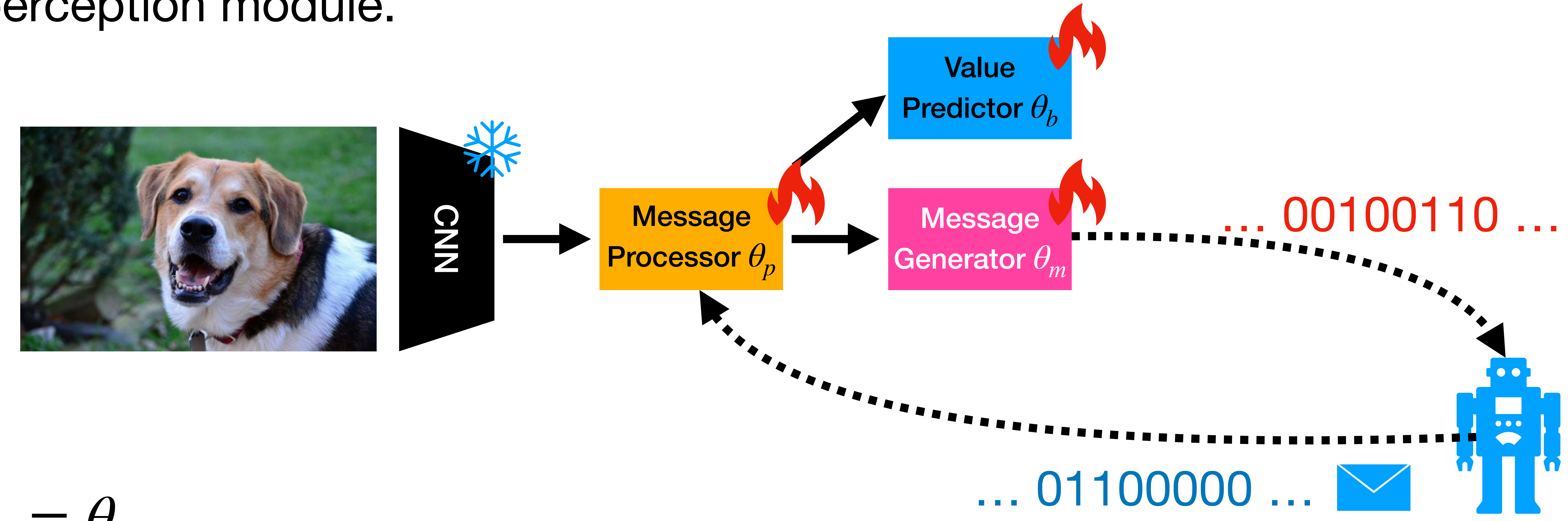
$$\theta_b \cap \theta_m = \theta_p$$

A minimalistic implementation

A Sender



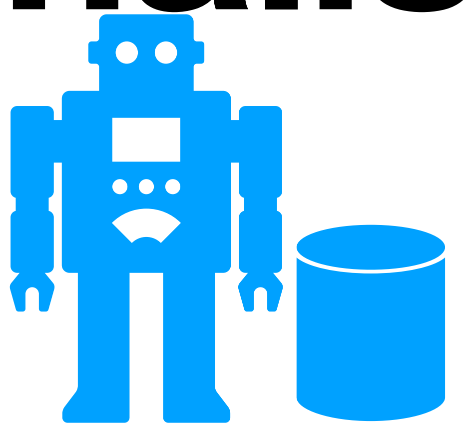
- The sender sees the image and answers a query from the receiver.
- It does not possess its own memory and is a very thin **wrapper** to the visual perception module.



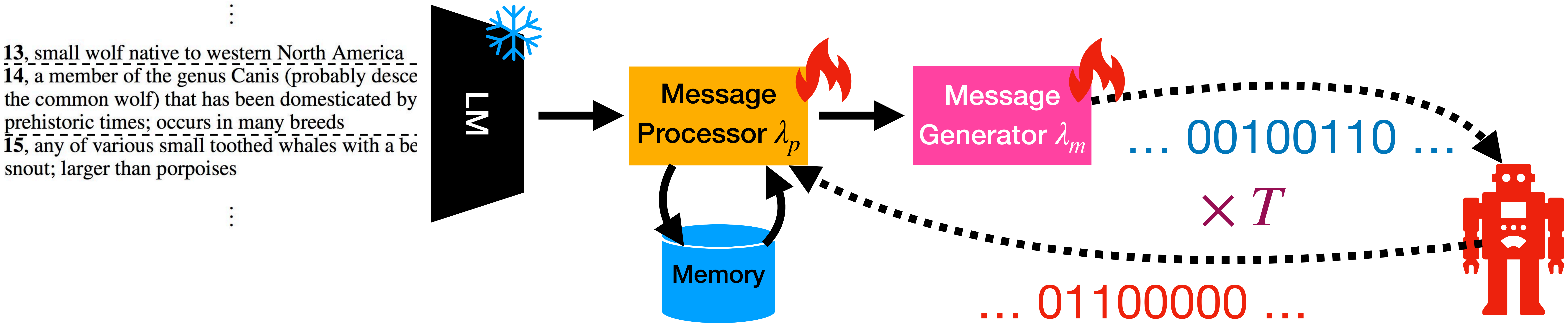
$$\theta_b \cap \theta_m = \theta_p$$

A minimalistic implementation

A receiver

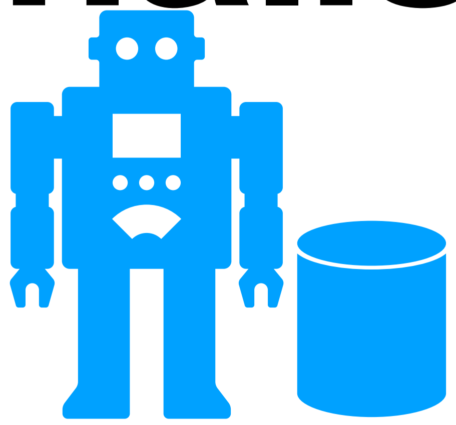


- The receiver uses a pretrained language model to read dictionary definitions.
- The receiver queries the sender for more information.
- The receiver has a memory to collect and accumulate evidence received from the sender over multiple turns of conversation.

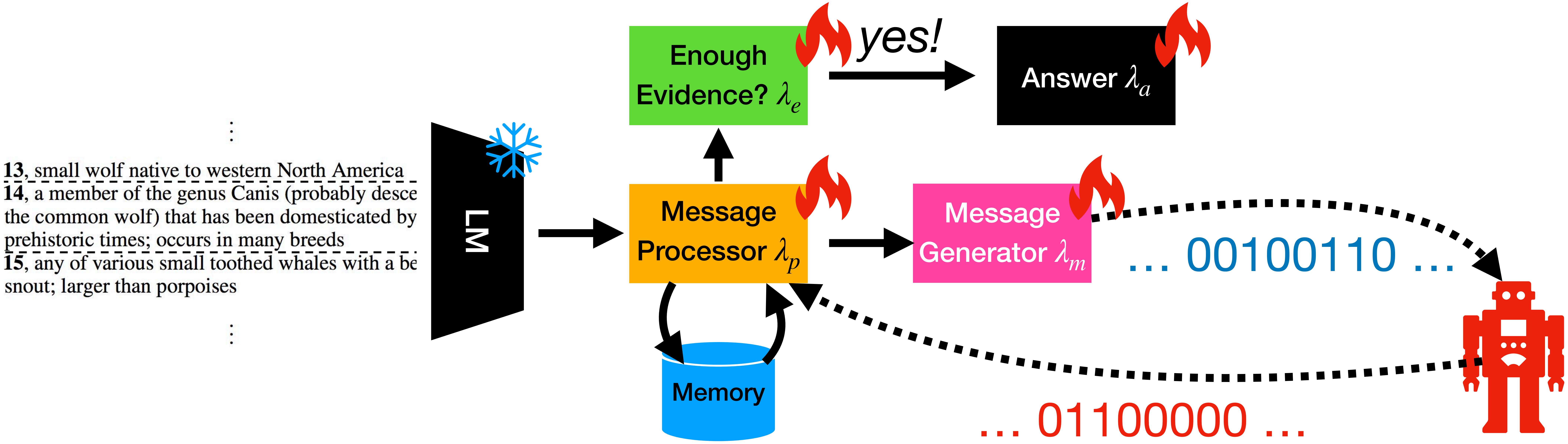


A minimalistic implementation

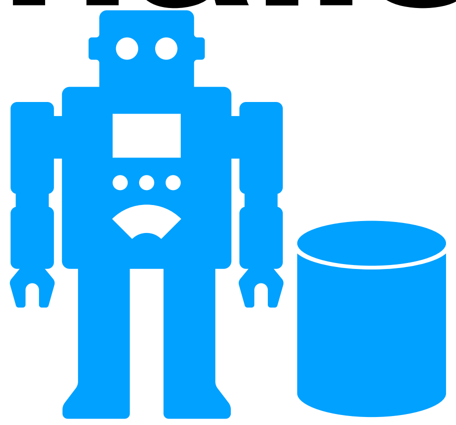
A receiver



- The receiver determines whether it has collected enough evidence.
- If so, it predicts which object the sender sees from the dictionary it reads.
- To simplify the setup, “answer” is done as classification over all definitions.

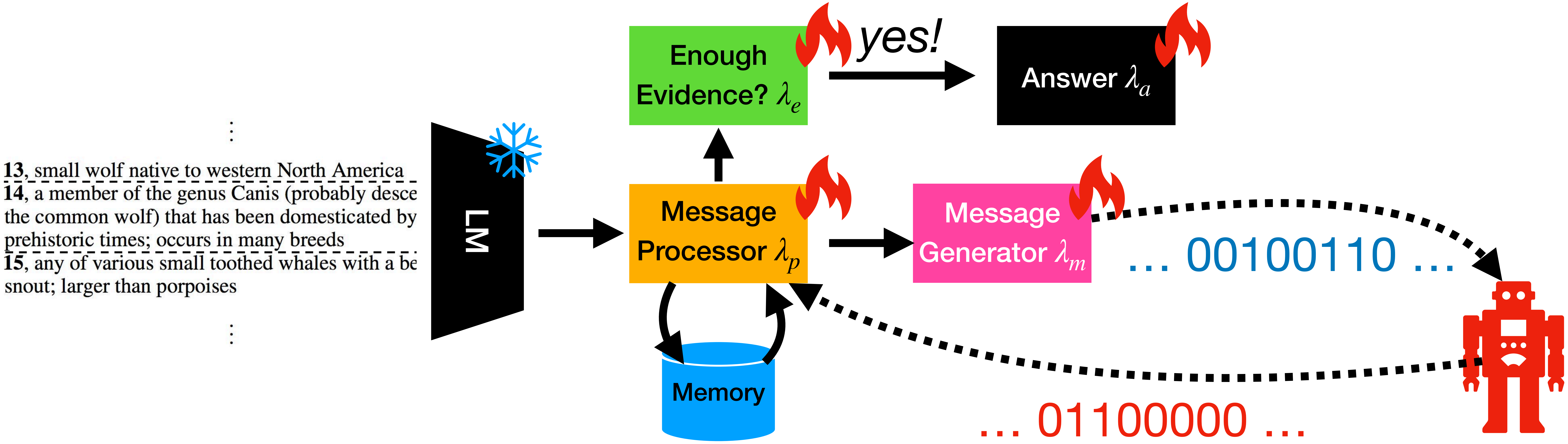


A minimalistic implementation

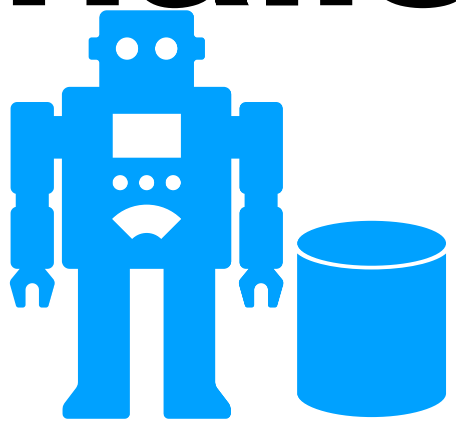
A receiver  : Steps

- The receiver updates the memory each time it receives a message from the sender

$$h^t \leftarrow M(h^{t-1}, \hat{m}^t, \mathrm{LM}(D))$$



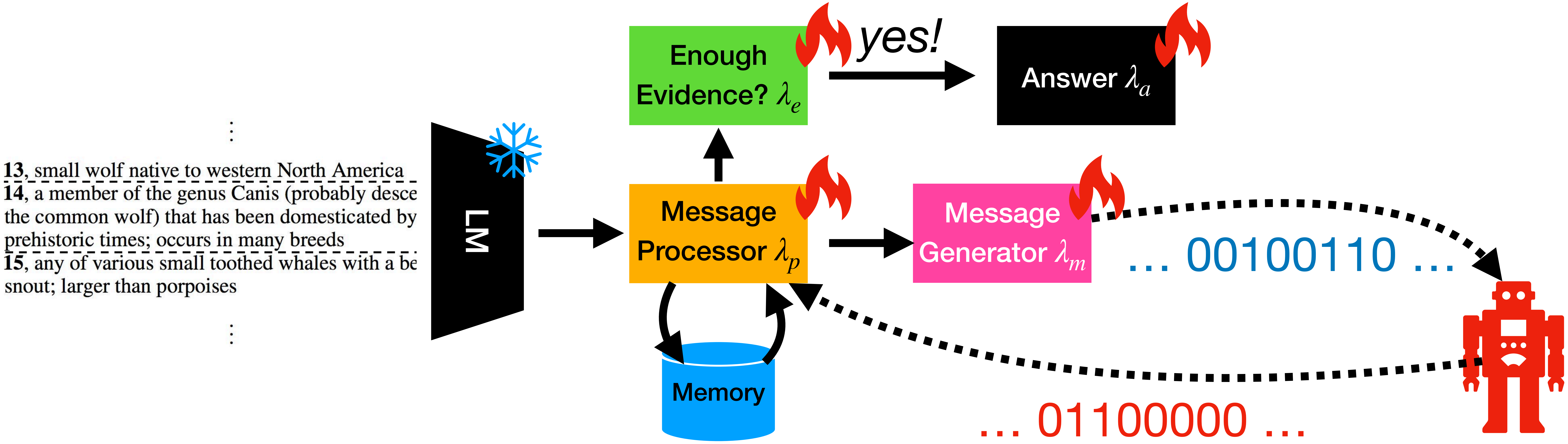
A minimalistic implementation

A receiver  : Steps

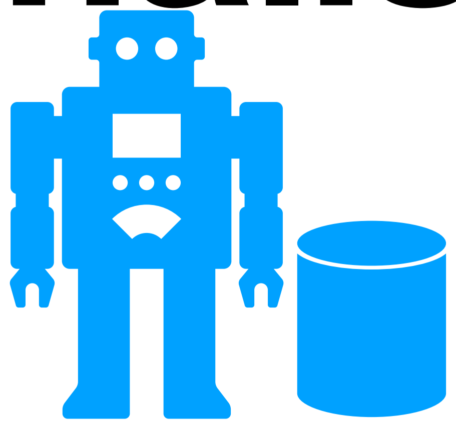
- The receiver determines whether enough evidence has been collected.

$$p(e^t \mid E(h^t))?$$

- This part requires RL for training.



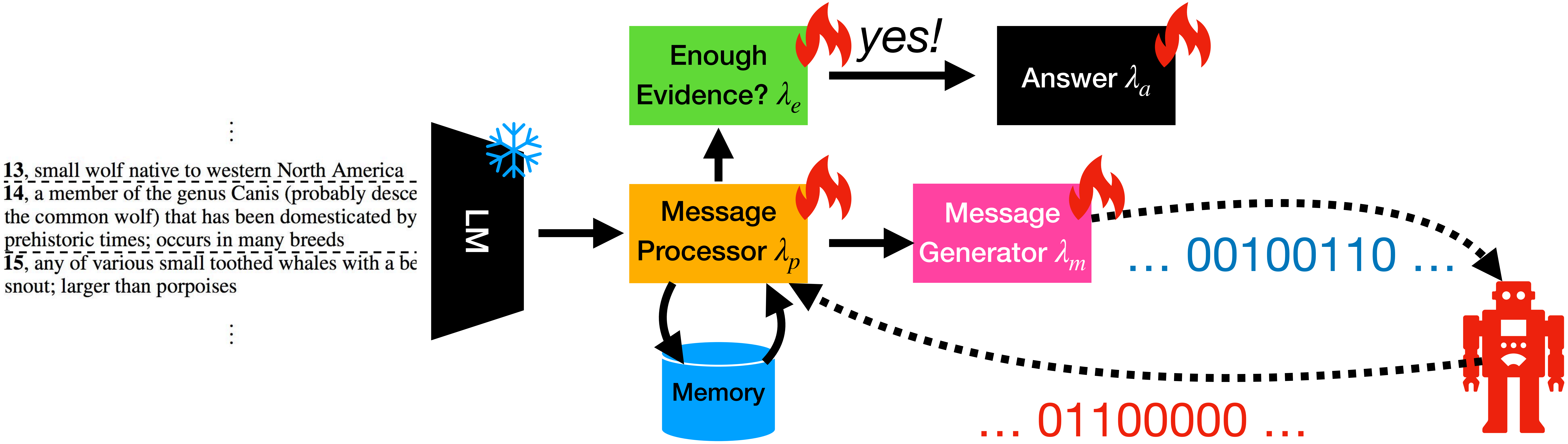
A minimalistic implementation

A receiver  : Steps

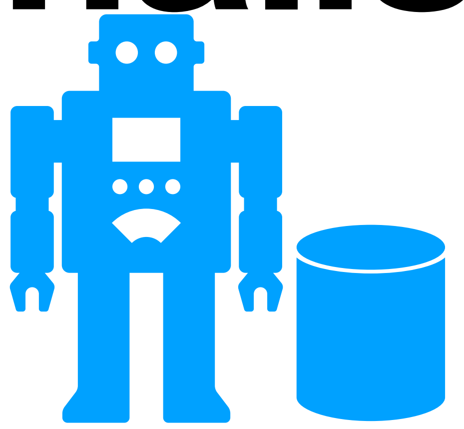
- If enough evidence, answer by choosing one of the definitions.

$$p(y^t \mid E(h^t), \mathrm{LM}(D))?$$

- We use supervised learning for this part in order to reduce variance in learning



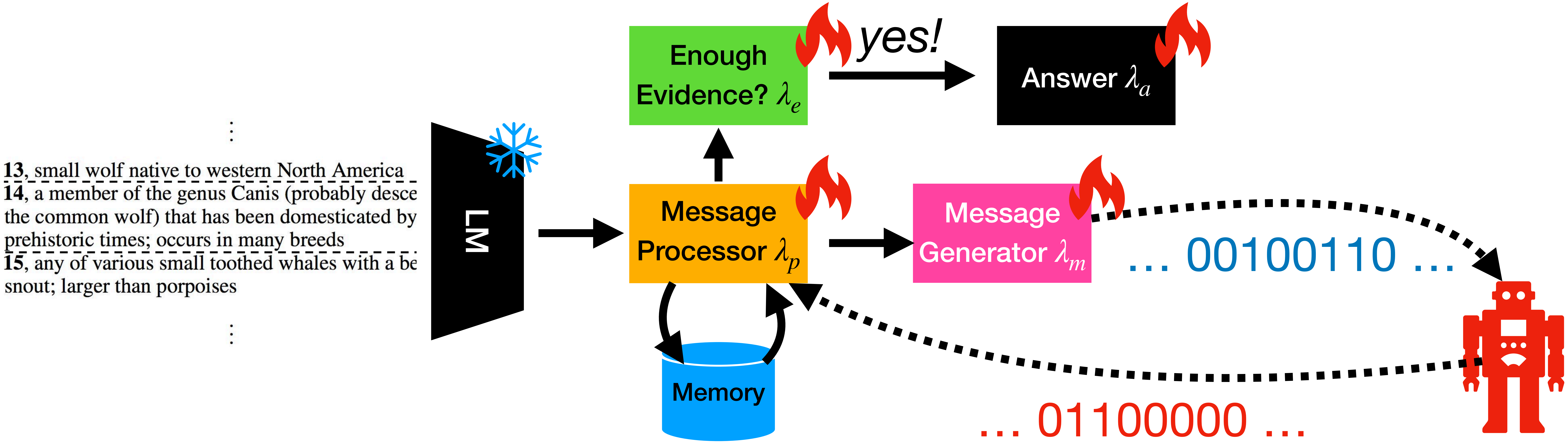
A minimalistic implementation

A receiver  : Steps

- If not enough evidence, query the sender with a new message.

$$p(m_i^t \mid h^t, \mathrm{LM}(D))$$

- Because messages are discrete, we use RL.



A minimalistic implementation

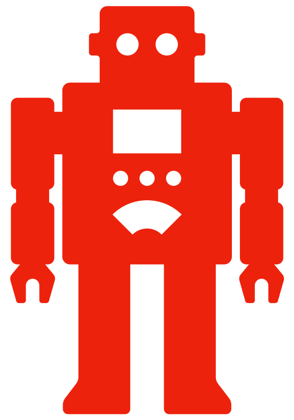
An overall end-to-end learning objective

$$L^i = L_c^i + L_r^i - \sum_{t=1}^{T} (\lambda_s H(s^t) + \lambda_m \sum_{j=1}^{d} (H(m_{s,j}^t) + H(m_{r,j}^t)))$$

So many terms, but all well justified and needed.

A minimalistic implementation

An overall end-to-end learning objective for the sender



$$L^i = L_c^i + L_B^i + L_r^i - \sum_{t=1}^{T} \left(\lambda_s H(s^t) + \lambda_m \sum_{j=1}^{d} \left(H(m_{s,j}^t) + H(m_{r,j}^t)\right)\right)$$

To improve exploration

$$L_B^i = \sum_{t=1}^{T} (R - B_s(o_s, m_r^{t-1}))^2 + (R - B_r(m_s^t, h_r^{t-1}))^2$$

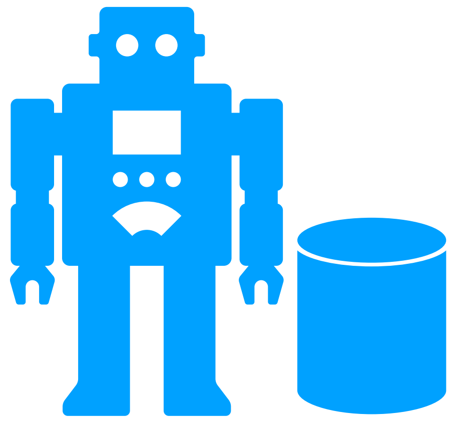
Baseline (value) prediction

$$L_r^i = \sum_{t=1}^{T} \left((R - B_s(o_s, m_r^{t-1})) \sum_{j=1}^{d} \log p(m_{s,j}^t) + (R - B_r(m_r^t, h_r^{t-1}))(\log p(s^t) + \sum_{j=1}^{d} \log p(m_{r,j}^t)) \right)$$

RL for learning to communicate

A minimalistic implementation

An overall end-to-end learning objective for the receiver



To improve exploration

$$L^i = L^i_c + L^i_B + L^i_r - \sum_{t=1}^{T} \left(\lambda_s H(s^t) + \lambda_m \sum_{j=1}^{d} \left(H(m^t_{s,j}) + H(m^t_{r,j})\right)\right)$$

Baseline (value) prediction

$$L^i_B = \sum_{t=1}^{T} (R - B_s(o_s, m_r^{t-1}))^2 + (R - B_r(m_s^t, h_r^{t-1}))^2$$

RL for learning to communicate

$$L^i_r = \sum_{t=1}^{T} \left((R - B_s(o_s, m_r^{t-1})) \sum_{j=1}^{d} \log p(m^t_{s,j}) + (R - B_r(m_r^t, h_r^{t-1}))(\log p(s^t) + \sum_{j=1}^{d} \log p(m^t_{r,j})) \right)$$

RL for learning to stop

$$L^i_c = - \log p(o^*_r = 1)$$

Classification loss

A minimalistic experiment

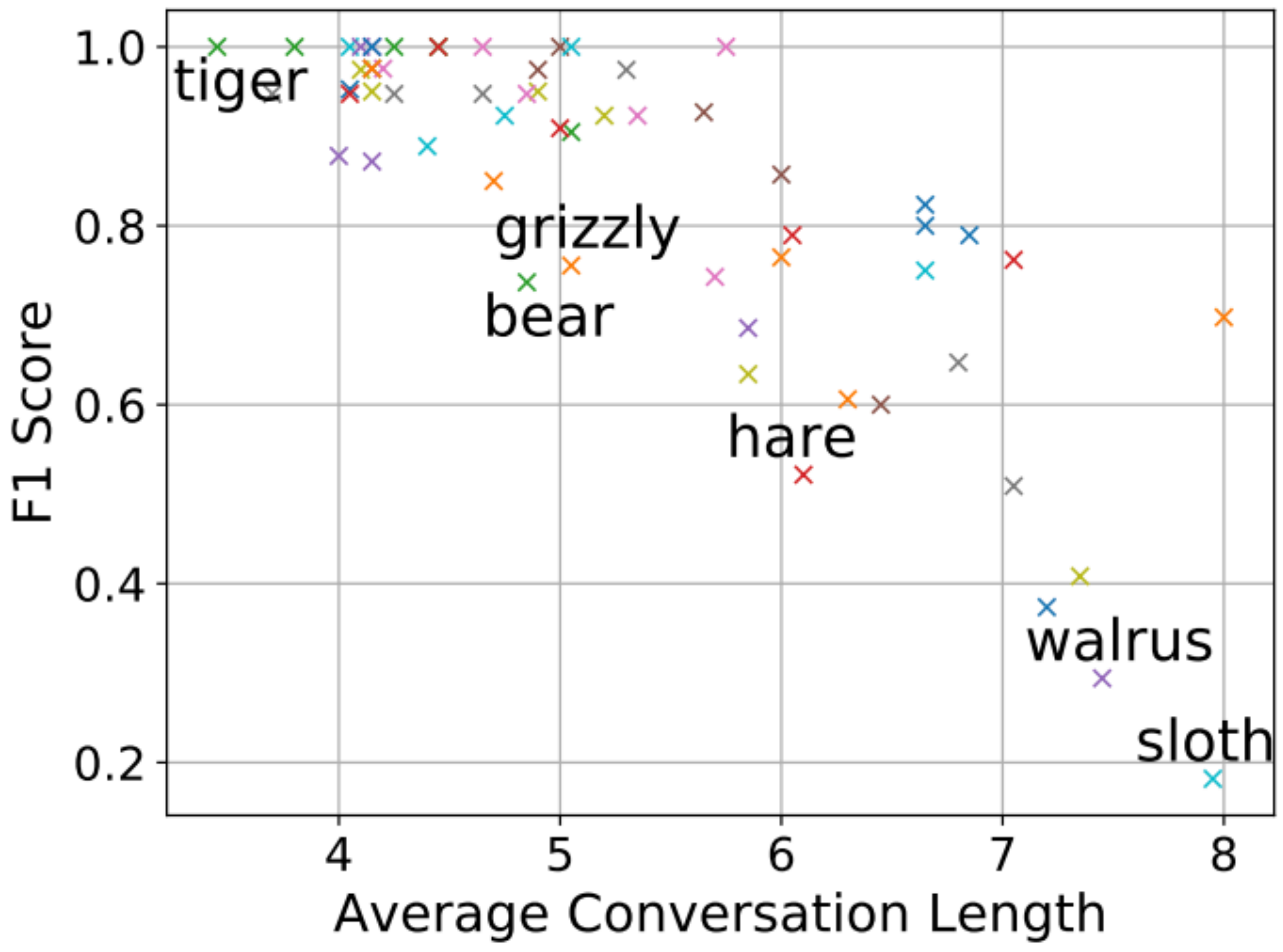
Guess a mammal!

- 60 mammals (for training and in-domain evaluation; 650 images each)
  + 10 mammals (for out-of-domain evaluation; 100 images each)
  + 10 insects (for transfer evaluation; 100 images each)
- Hyperparameter tuning on the in-domain validation examples.
- Training with RMSProp [Tieleman & Hinton, 2012] for 8 hours an experiment.
- Max. # of turns capped at 10 during training.
- Evaluation metric: Accuracy@K with K=10% of the number of categories.
- All implemented using PyTorch.

Analyses

Test-time compute vs. difficulty

- The length of the conversation correlates strongly with the difficulty of the categories.
- The difficulty is measured by the category's accuracy (F1) on a *separate* classifier.
- The agents know to use a longer conversation to resolve ambiguity when they are dealing with a more challenging problem.

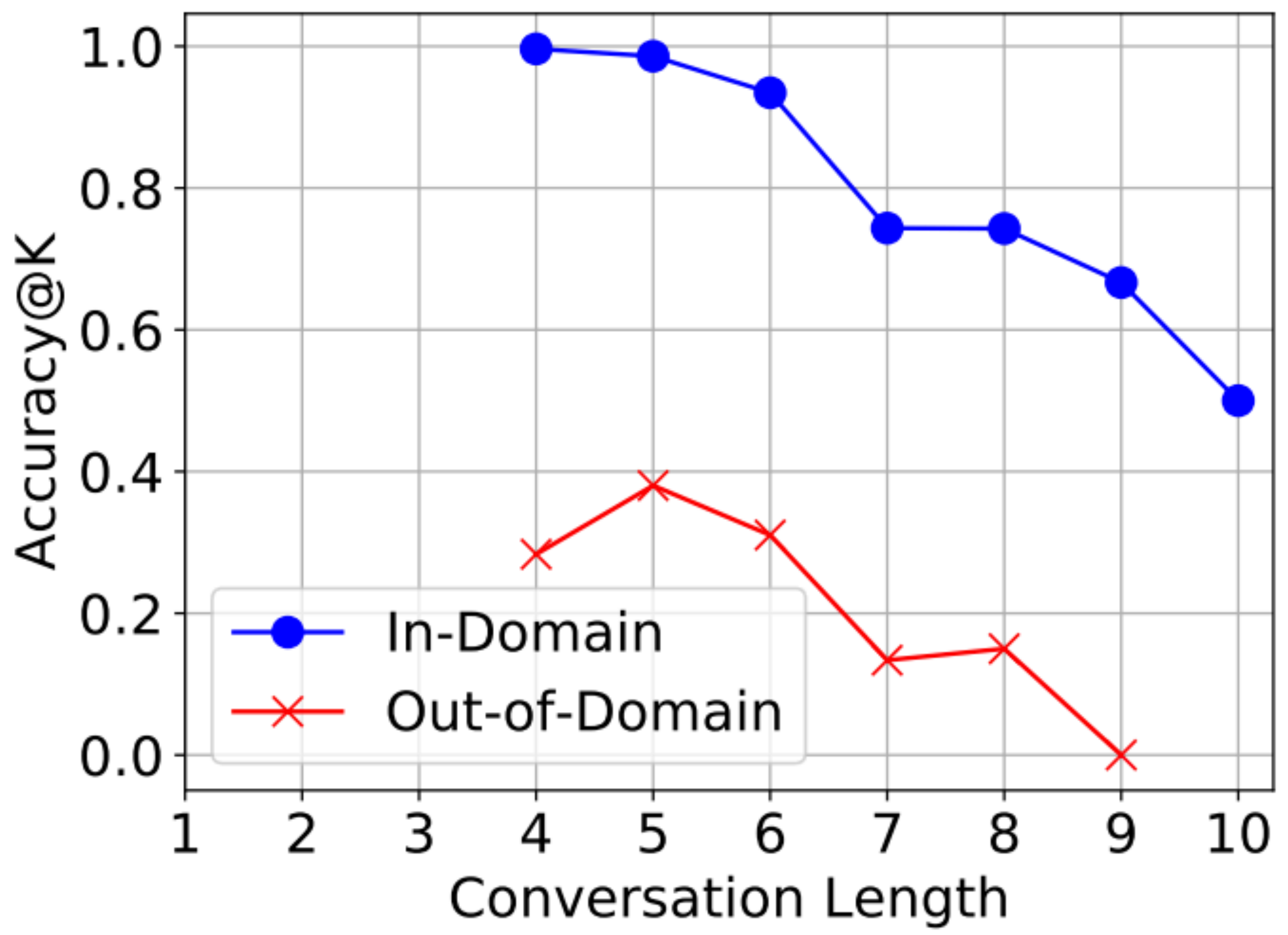


(a) Difficulty

Analyses

Test-time compute vs. difficulty

- The agent’s accuracies also reveal this relationship: the longer the conversation was the lower the accuracy ends up being.
- This trend persists whether an instance is in-domain or out-of-domain.
- This behavior emerges naturally from learning; the receiver learns to stop when enough evidence is collected.

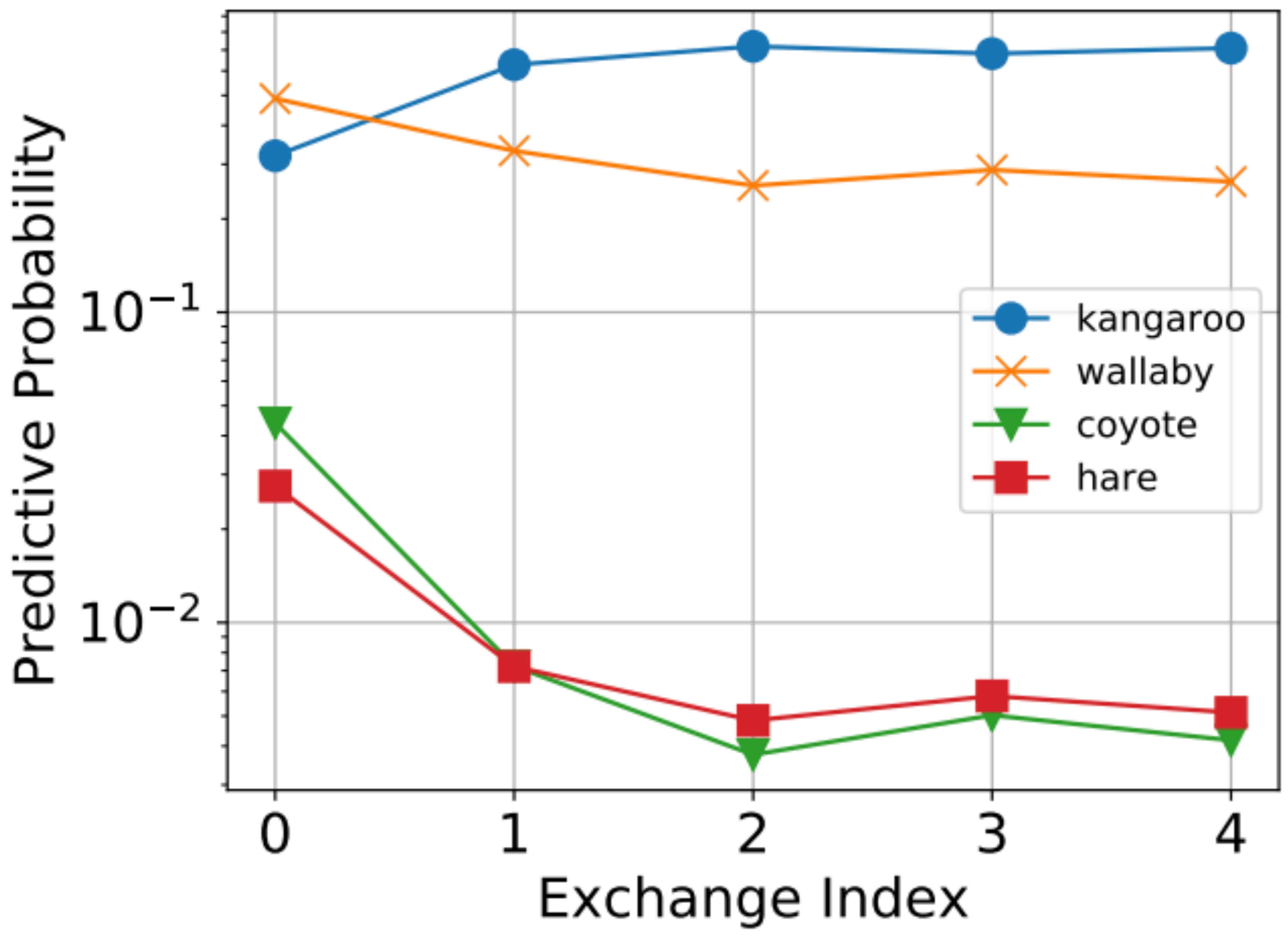


(b) Accuracy

Analyses

Resolving ambiguity in the test time over the conversation

- At the very first turn (0), the probabilities of categories are clustered with each other.
- In particular, the receiver thinks the image may be of either “kangaroo” or “wallaby”; indeed confusing!
- As the conversation continues, the agents collectively resolve this ambiguity; predicting “kangaroo”!

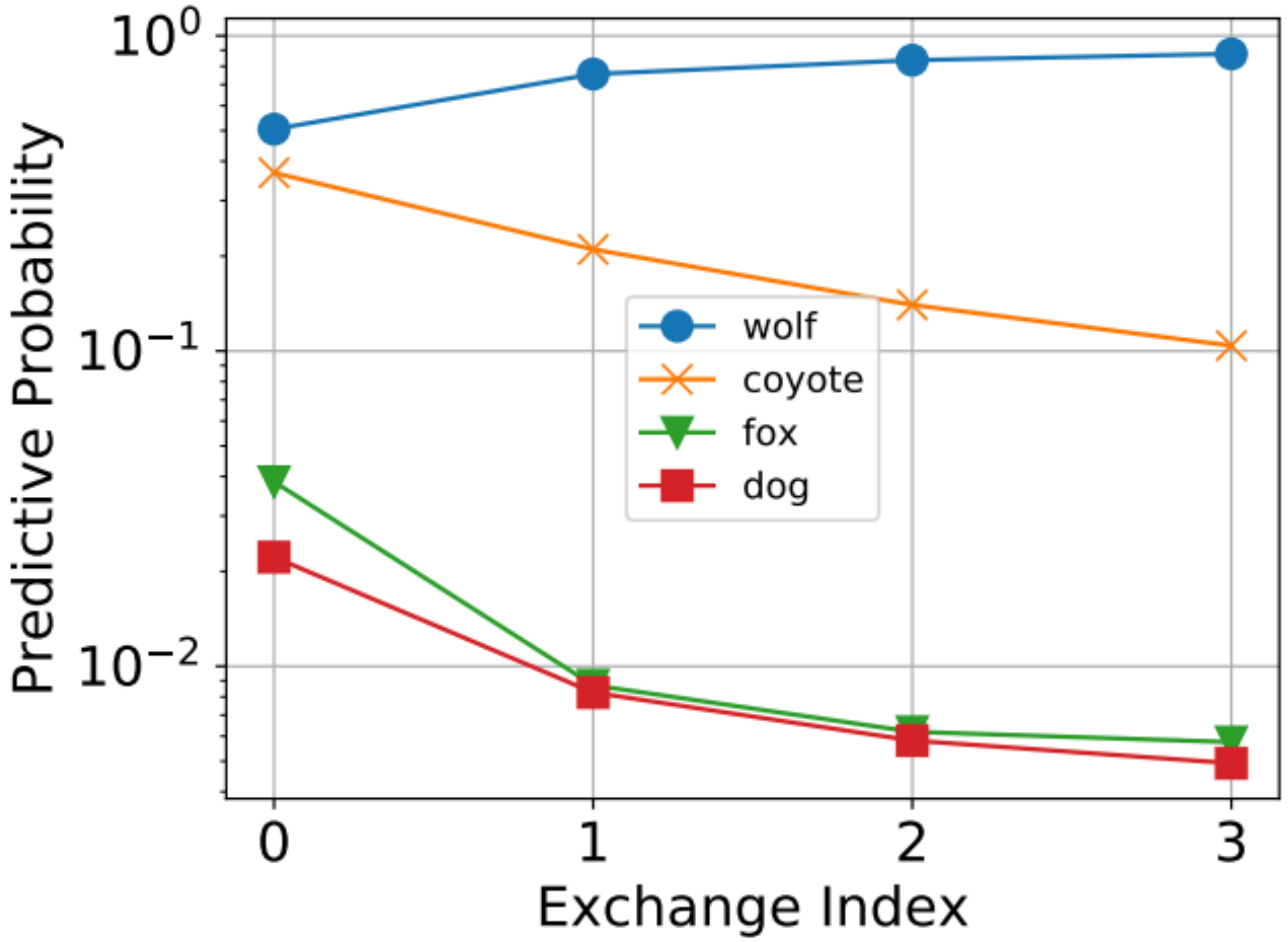


(b) Kangaroo

Analyses

Resolving ambiguity in the test time over the conversation

- Not an isolated case, but a general one; the agents know how to communicate with each other to solve a problem collectively.

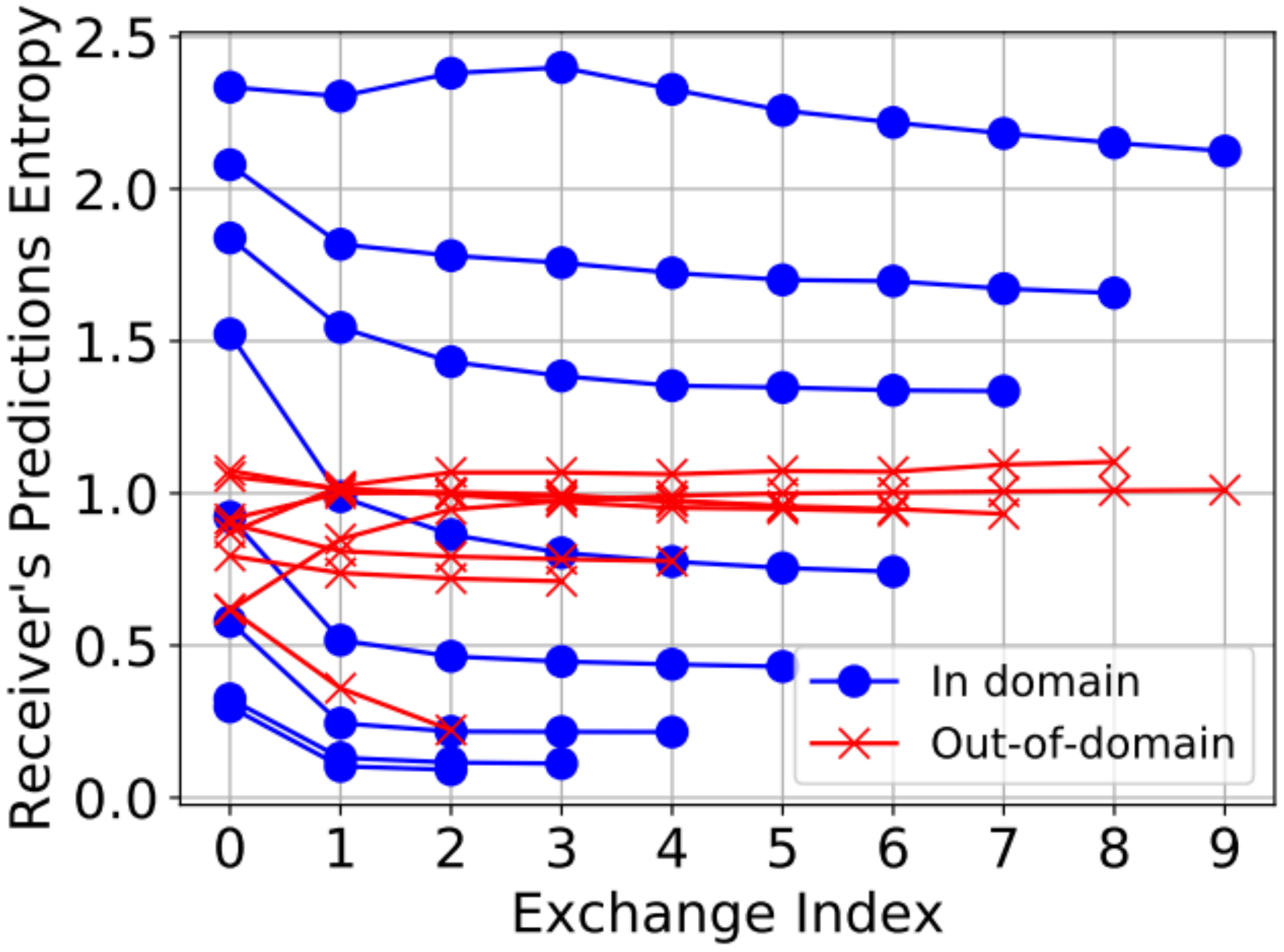


(c) Wolf

Analyses

Resolving ambiguity in the test time over the conversation

- This can be measured more explicitly by the predictive entropy of the receiver.
- As more evidence is received from the sender, the receiver is increasingly more confident about its prediction.
- This trend is less so with out-of-domain instances, implying that domains matter even for learning to be confident (or uncertain).



(a) Predictions

Analyses

Information theoretic analysis

- As more evidence is collected, the receiver is increasingly more confident about what to ask further. This is reflected by the decreasing entropy of the message distribution over the conversation.
- On the other hand, the sender is asked increasingly more challenging queries, resulting in less confidence in its answers. This is reflected on the increasing entropy over the conversation.

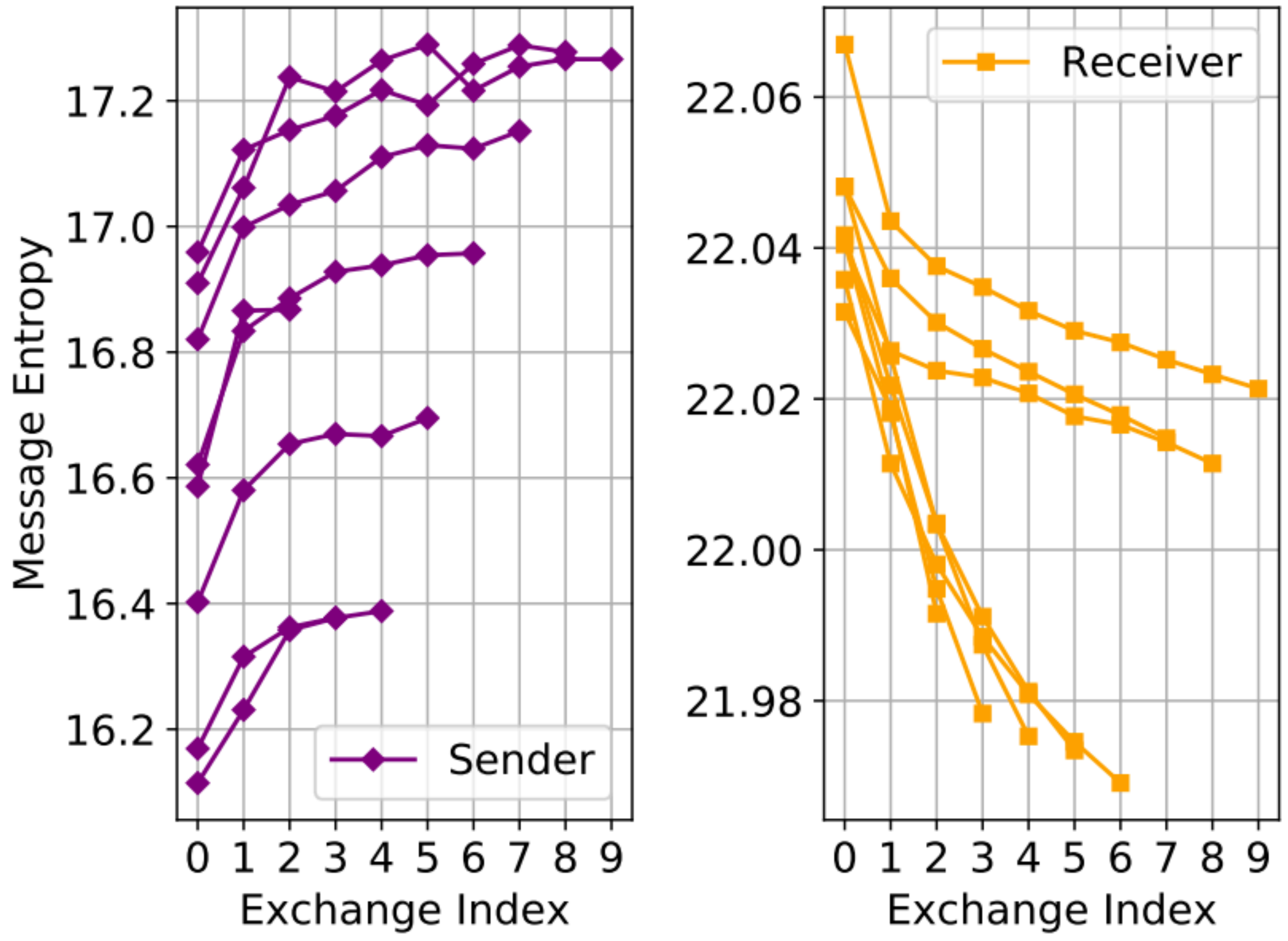


Figure 4: Message entropy over the conversation on the in-domain test set of the sender (left) and receiver (right).

Analyses

Is end-to-end learning necessary?

- Do the agents need to adapt to each other in order to achieve the joint goal better? What if only the receiver learns to work with a fixed sender?
- We observe that end-to-end learning always exhibits a significant improvement over receiver-only learning.
- This suggests the importance of considering the overall agentic system end-to-end, in order to make the most out of it.

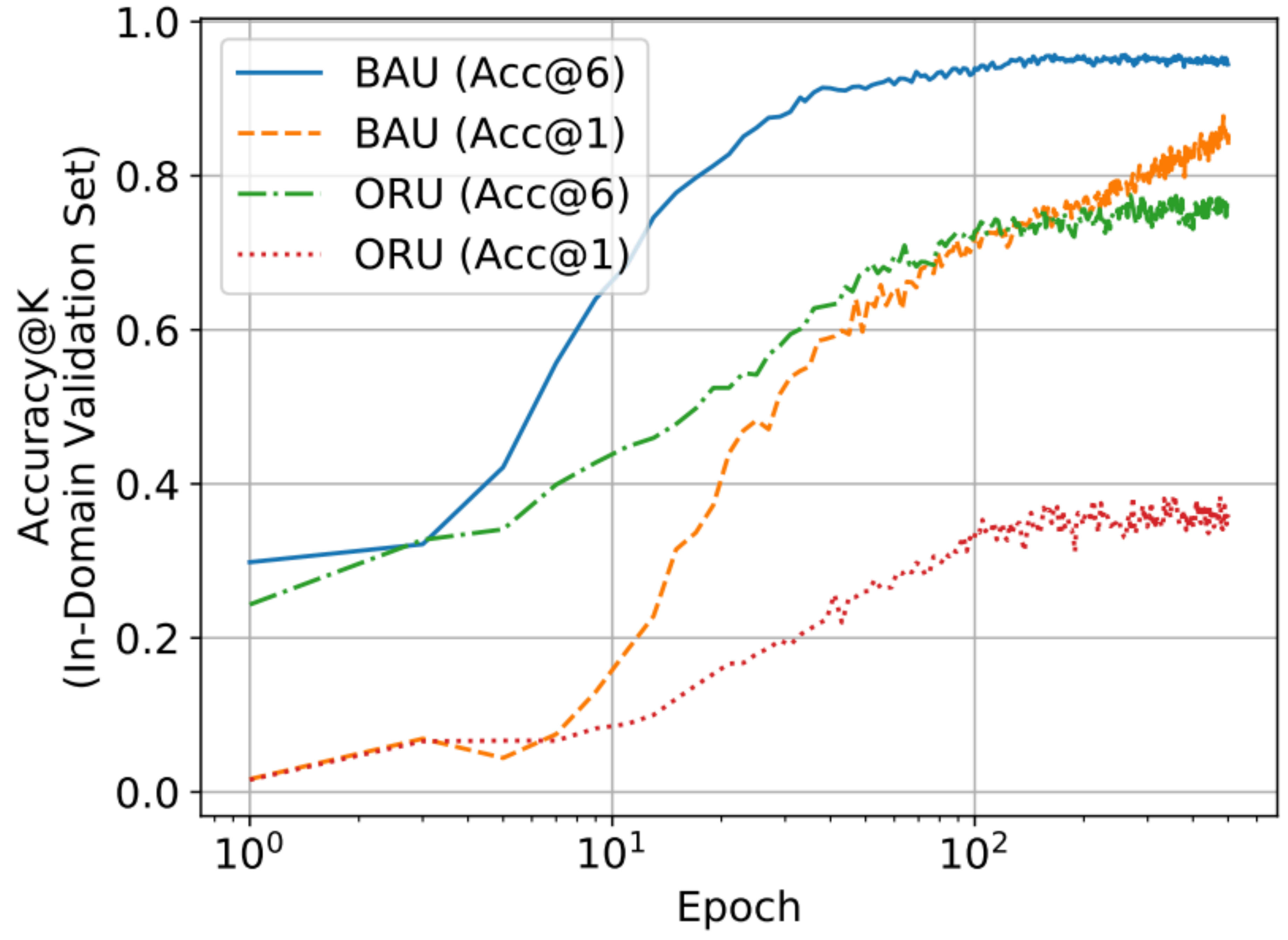
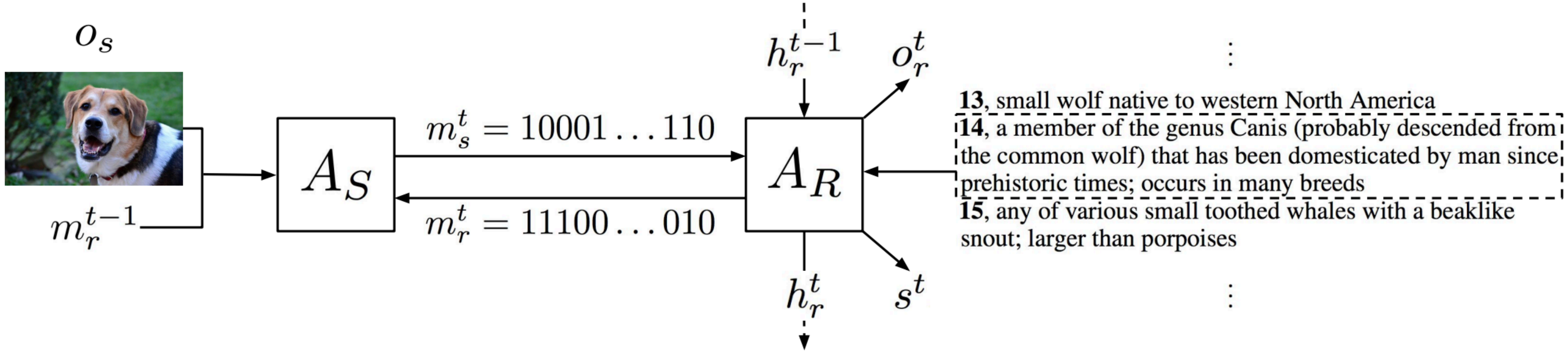


Figure 6: Learning curves when both agents are updated (BAU), and only the receiver is updated (ORU).

A multi-modal, multi-turn, multi-agent system

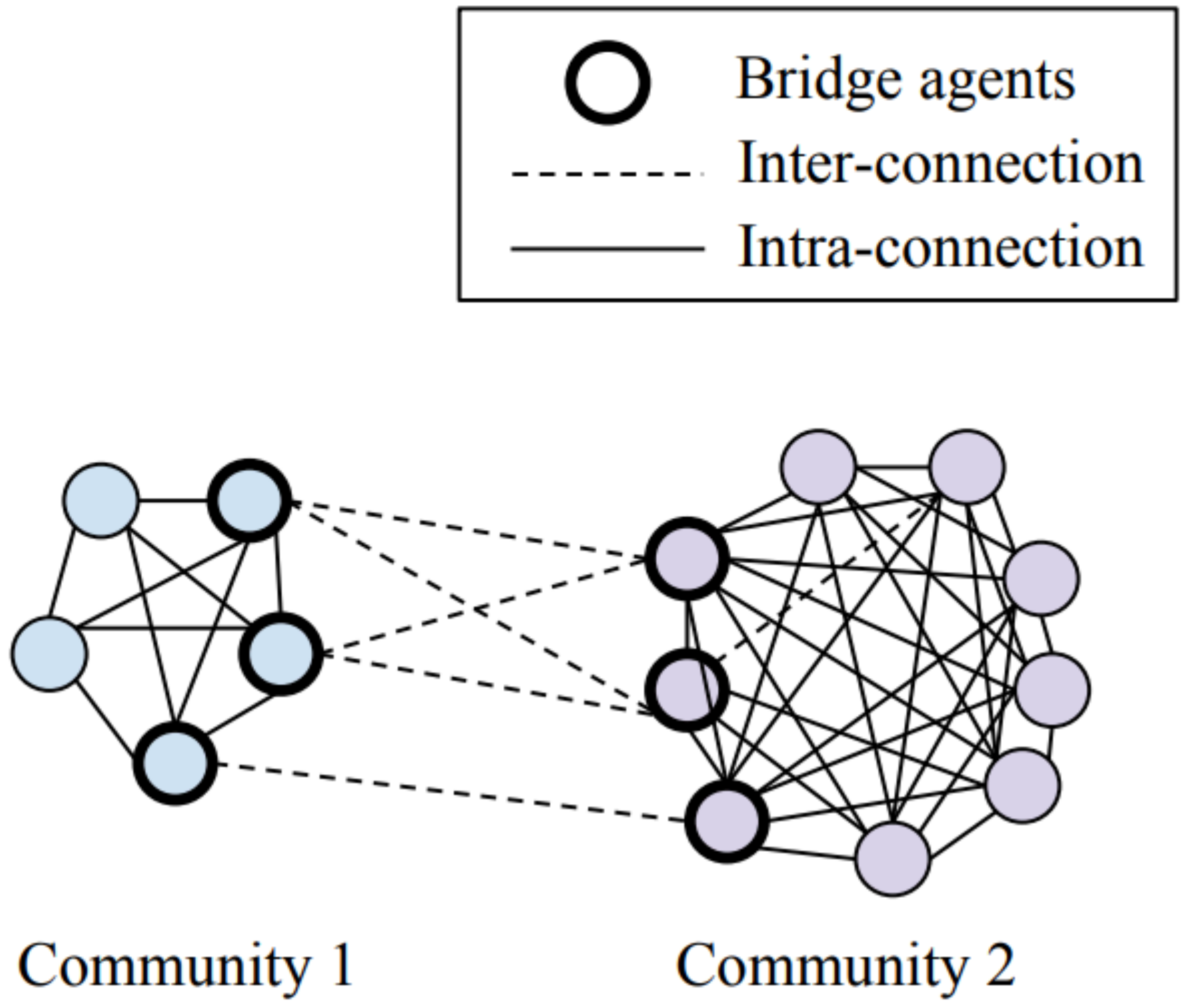
A quick summary

- Two agents play a referential game in a multi-turn conversation.
- These agents are not symmetric (modality and the existence of memory).
- Ambiguities are resolved adaptively over a dynamic-length conversation.
- These agents are trained end-to-end using a hybrid supervised-reinforcement learning strategy, to maximize the joint reward.



A multi-modal, multi-turn, multi-agent system

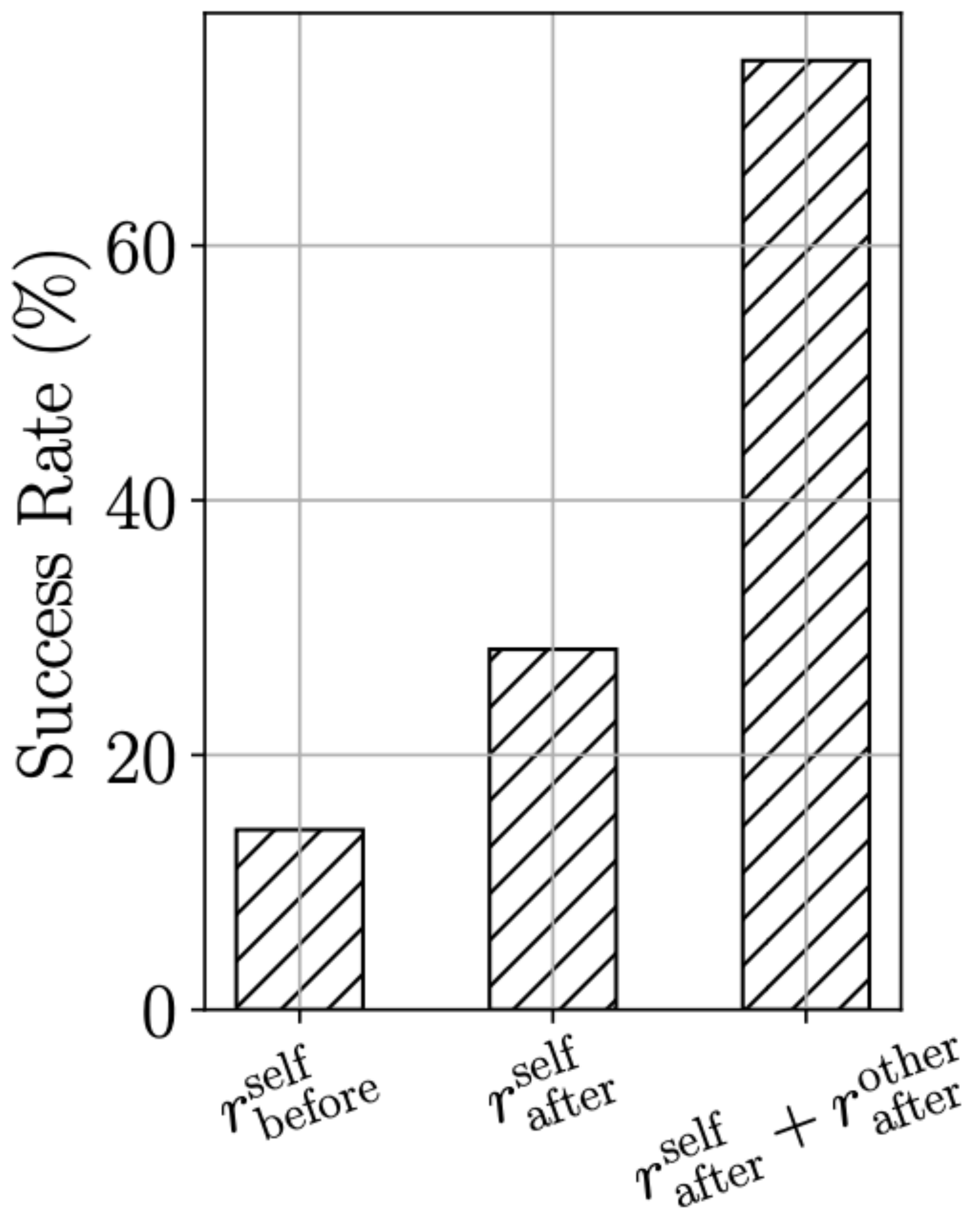
Future directions



- This work was extended to a setting with many agents with non-trivial topologies [Graesser et al.]. In this extension, it was also shown that a shared, joint reward is needed to encourage agents to learn to cooperate.

- More studies in a more realistic setup are being done and will be important, in order to establish the utility of multi-agent systems.

I don't think she's on a job market.



fin.

So, where are these papers?

Published as a conference paper at ICLR 2018

Emergent Communication in a Multi-Modal, Multi-Step Referential Game

Katrina Evtimova[1], **Andrew Drozdov**[2], **Douwe Kiela**[3], and **Kyunghyun Cho**[1,2,3,4]

[1]Center for Data Science. New York University
[2]Department of Computer Science. New York University
[3]Facebook AI Research
[4]CIFAR Azrieli Global Scholar

Abstract

Inspired by previous work on emergent communication in referential games, we propose a novel multi-modal, multi-step referential game, where the sender and receiver have access to distinct modalities of an object, and their information exchange is bidirectional and of arbitrary duration. The multi-modal multi-step setting allows agents to develop an internal communication significantly closer to natural language, in that they share a single set of messages, and that the length of the conversation may vary according to the difficulty of the task. We examine these properties empirically using a dataset consisting of images and textual descriptions of mammals, where the agents are tasked with identifying the correct object. Our experiments indicate that a robust and efficient communication protocol emerges, where gradual information exchange informs better predictions and higher communication bandwidth improves generalization.

[Evtimova et al., 2018 ICLR]

Emergent Linguistic Phenomena in Multi-Agent Communication Games

Laura Graesser[†*], **Kyunghyun Cho**[†‡⋆], **Douwe Kiela**[‡]
[†] NYU; [*] Robotics at Google; [⋆] CIFAR Azrieli Global Scholar; [‡] Facebook AI Research
lauragraesser@google.com, kyunghyun.cho@nyu.edu, dkiela@fb.com

Abstract

We describe a multi-agent communication framework for examining high-level linguistic phenomena at the community-level. We demonstrate that complex linguistic behavior observed in natural language can be reproduced in this simple setting: i) the outcome of contact between communities is a function of inter- and intra-group connectivity; ii) linguistic contact either converges to the majority protocol, or in balanced cases leads to novel creole languages of lower complexity; and iii) a linguistic continuum emerges where neighboring languages are more mutually intelligible than farther removed languages. We conclude that at least some of the intricate properties of language evolution need not depend on complex evolved linguistic capabilities, but can emerge from simple social exchanges between perceptually-enabled agents playing communication games.

where agents are neural networks endowed with the ability to exchange messages about their perceptual input. The advantage of this approach is that one can precisely control linguistic, environmental and algorithmic variables.

First, we investigate linguistic behavior at the agent-level, and examine when symmetric communication emerges within a linguistic community, as well as how the topological organization of communities—i.e., which other agents an agent comes into contact with and how frequently that happens—impacts convergence and learning. We then examine the behavior of communities of such agents when they come into contact, as well as how community-level topology impacts convergence, success rate and mutual intelligibility.

We demonstrate that the following linguistic b haviors emerge, which correspond to known li guistic phenomena in natural languages: 1) t outcome of contact is a function of inter- a

[Graesser et al., 2019 EMNLP]