Learning how, what, and whether to communicate: emergence of protocommunication in reinforcement learning agents

Takashi Sato stakashi@oist.jp
Eiji Uchibe uchibe@oist.jp
Kenji Doya doya@oist.jp
Initial Research Project (IRP), Okinawa Institute of Science and Technology (OIST)
12-22, Suzaki, Uruma, Okinawa, Japan, 904-2234

Abstract

This paper examines whether and how a primitive form of communication emerges between adaptive agents by using their excess degrees of freedom in action and perception. As a case study, we consider a game in which two reinforcement learning agents learn to earn rewards by intruding into the other's territory. Our simulation showed that the agents with lights and light sensors could learn turn-taking behaviors by avoiding collisions using visual communication. Further analysis revealed that there was a variety in what message is mapped to what signal, and in some cases there was role differentiation into a sender and a receiver.

Keywords: Reinforcement Learning, Intrusion Game, Emergence of Protocommunication, Role Differentiation

1 Introduction

The prototype of communication, or protocommunication, would have emerged to help individuals to earn rewards and to improve fitness. Then how did protocommunication emerge based on what capacity of individuals? These questions have been discussed in various fields for a long time, but unlike other questions in archeology, these questions are hard to answer as there is no fossil of communication until written languages emerged recently. Thus, we address this question by "understanding by construction" using mathematical modeling and computer simulations [1].

The studies on emergence of communication can be classified into two broad categories: one adopting evolutionary optimization [2, 3] and the other employing learning agents [4, 5]. However, a major limitation in most previous studies is that they assumed the preexistence of the basic frameworks for communication, such as signals and meanings or a speaker and a hearer, and just verified the evolution or learning of mappings between signals and meanings by taking the success of communication itself as the objective function. Therefore, it is difficult for those studies to answer how communication emerged from a world where concepts like signals, words, and speaking did not exist.

The purpose of this study is to test if communication can emerge between individuals who have basic behavior learning functions but do not have dedicated mechanisms or absolute needs for communication. Specifically, we run a case study of an "intrusion game" in which two agents move on a linear track and earn rewards by intruding into the other's territory while avoiding collisions. We consider what action, sensation and memory capacities are necessary for learning of cooperative behaviors, and when it is learned, what meanings agents assign to their excess degrees of action and sensation. Further, we investigate the developmental process of cooperative behaviors by communication and the cases of role differentiation into a speaker and a hearer.

2 Intrusion Game

We consider an "intrusion game (IG)," which simplifies situations like a turf war between foraging animals. Two players can move back and forth on a one dimensional space with four slots. Players are bounded by walls on the "west" and "east" ends of the track and cannot jump over or stay together in the same slot with another agent. Figure 1 depicts six possible sets of positions that the players can takes. We denote the six position patterns by 0 to 5. The "west" player can get a reward by entering the east half of the track (i.e., position pattern 5) and the "east" player by entering in the west half (i.e., position patter 2) without a collision. A punishment (negative reward) is given when a player collides with a wall or another player.

A crucial problem in this game is how the players resolve the conflict at the position pattern 4. If the players act selfishly, i.e., to maximize its own reward,

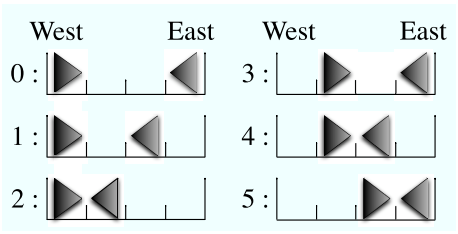

Figure 1: Six possible position patterns of the players, denoted by 0 to 5.

both would take an action to move forward, but it will cause a collision with negative rewards to both players.

3 Reinforcement Learning Agents

In order to test whether agents with general action learning capability can also learn to communicate, we adopt reinforcement learning agents [6] which can learn various behaviors based on rewards and punishment. We use the Q-learning method [6] which is standard for discrete tasks like IG. Q-values are updated by the following equation.

$$Q(s_t, a_t) := Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right] ,$$

where α is a learning rate ($0 < \alpha < 1$), γ is a discount rate ($0 < \gamma < 1$), and $r(t)$ is the reward given after action $a(t)$ was taken at state $s(t)$. We used the ϵ-greedy policy in which an actions is randomly selected with probability ϵ and otherwise an action that maximize Q-value for a given state is selected.

In order to investigate how the agents' sensory, action, and memory capabilities affect the learned behaviors, we tested four types of agents. A null or N-type agent simply has two moving actions (backward or forward) and can sense the position pattern (0 to 5) of the two agents. In addition, a light-capable, or L-type agent has actions of turning on or off its headlight and also a light sensor to see if the other agent's light is on or off. A memory-based, or M-type agent keeps the memory of its previous action (backward or forward) to augment its state space. A light-and-memory, or LM-type agent has both light signaling and memory capabilities.

4 Simulation Results

We performed 10 simulation runs each for the four types of agents with the following setups: positive reward +1 for successful intrusion, negative rewards -1 each for collision with the wall and the other agent, $\epsilon = 0.01$, $\alpha = 0.01$, and $\gamma = 0.9$.

4.1 Agents' Behaviors

First, we present examples of typical behavioral patterns obtained from the analysis of the change of each agent's position pattern (Fig. 2).

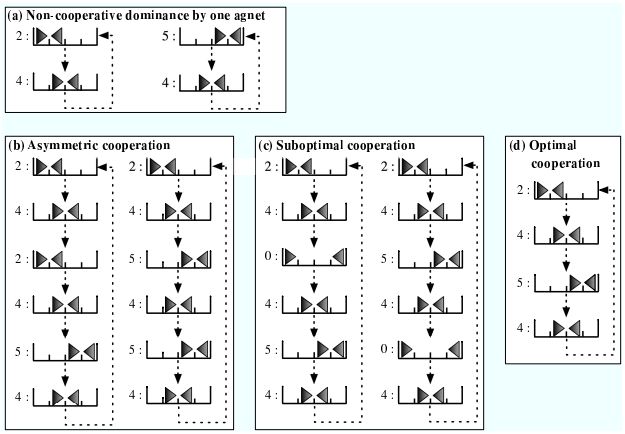

Figure 2: Four examples of typical behavioral patterns. (a) Non-cooperative dominance by one agent. (b) Asymmetric cooperation. (c) Suboptimal cooperation. (d) Optimal cooperation.

Figure 2(a) shows a non-cooperative dominance by one agent. In this pattern, only one agent can earn a positive reward every two steps. The other can get no reward, but can receive negative reward if it changes its behavior. Figure 2(b) presents an asymmetric cooperation which can be seen only in LM-type agents. In this case, one agent can get two positive rewards during a six step cycle, the other can obtain while only once. Figure 2(c) depicts a suboptimal cooperation leading to one reward every six steps. Figure 2(d) shows an optimal cooperation in which both agents earn a reward every four steps.

We analyze the occurrence frequency of four typical behavioral patterns for the four types of agents. As can be seen in Fig. 4, the agents without light (N- and M-type) can learn only the non-cooperative dominance. In contrast, the agents with light (L- and LM-type) can show various cooperation. Further, LM-type agents can achieve the optimal cooperation more frequently than L-type agents without action history.

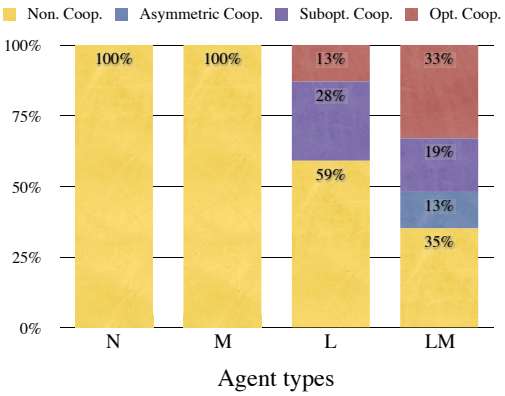

Figure 3: Occurrence frequency of four typical behavioral patterns for four types of agents.

4.2 Developmental Process

Next we examined how the behaviors changed by learning before converging to one of four typical patterns. Figure 4 shows the developmental history of four LM-type agents who acquired one of four typical behavioral patterns at final episode. We recorded the Q-values of the agents every 1,000 steps and let the agents play the IG (with $\epsilon = \alpha = 0$) from all possible initial states[1].

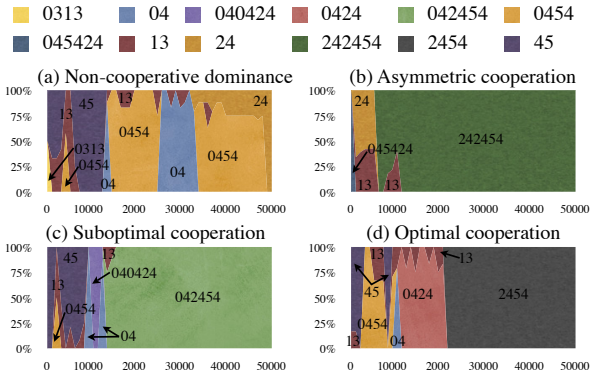

Figure 4: History of occurrence frequency of behavioral patterns. The x-axis and the y-axis of each figure are the steps and the occurrence frequency of converged behavioral patterns, respectively. Each numerical string represents a sequence of position patterns in a cyclic behavior.

Figure 4(a) shows an example of the history of a pair that converged to non-cooperative dominance by the east agent. The other diagrams in Fig. 4 indicate (b) asymmetric, (c) suboptimal and (d) optimal cooperation. A common feature in these cooperative cases that the agents experienced took both position pattern sequences 4 to 2 and 4 to 5 in the early stage before becoming able to switch between the two.

[1]For example, the number of states of the L-type agents is 24, where initial positions and light states are 6 and 4, respectively.

4.3 Variety of Signaling

We observed emergence of various types of communication emergence. Figure 5 exemplifies four typical types of communication that realizes the optimal cooperation.

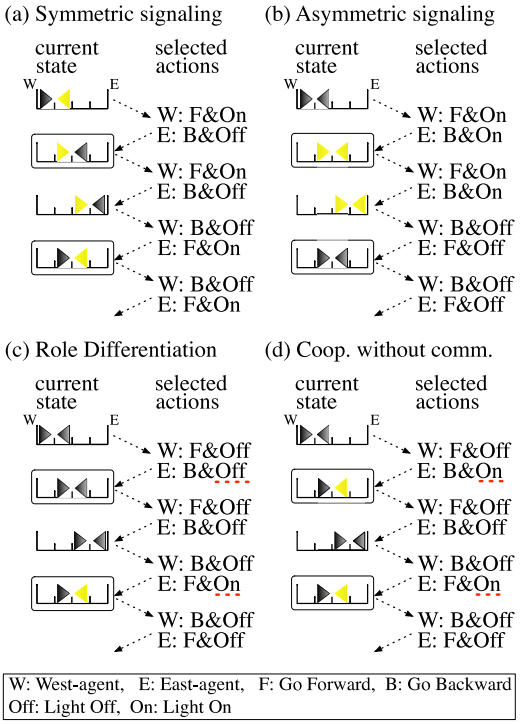

Figure 5: Typical examples of emerged communication. (a) Symmetric signaling. (b) Asymmetric signaling. (c) One-way communication between a sender and a receiver after a role differentiation. (d) Cooperation without communication.

Figure 5(a) shows an example of symmetric signaling in which agents can resolve the conflict at the position pattern 4 by alternately turning on the light while stepping forward. This means that the agents can convey their next actions as messages by their lights. An asymmetric signaling can also be observed, in Fig. 5(b), in which one agent turns the light on to step forward, while another agent turns the light on to step back.

Unlike the symmetric and asymmetric signaling, communication examples shown in Fig. 5(c) and (d) can be seen only in LM-type agents. Figure 5(c) depicts a one-way communication between a sender (the east agent) and a receiver (the west agent) after a role differentiation. In this case, only the east agent uses its light source, and the west agent behaves according to the east agent's light signal. The east agent, who cannot rely on the light signal of the west agent, determines its actions based on memory of its own past action in order to solve the conflict on the position pattern 4. Figure 5(d) is a case of optimal cooperation without communication. As can be seen, the east agent always turns on its light when entering the position pattern 4 from 2 or 5. Therefore, the west agent cannot solve the conflict on the position pattern 4 by using the east agent's light signal. Both agents behave based only on the memory of their past actions.

5 Discussion

Our simulation study showed that simple communication can emerge from iteration of searching for roles of redundant actions through a generic reinforcement learning process and interaction with the others.

Animal communication is defined as a transmission of signals that senders can profit by reactions of receivers [7]. Our simulation confirms that this type of communication can be emerged from repeated interactions between reinforcement learning agents with enough physical capacity. Tomasello claims that communicative signals can be created by forming each other's behaviors between two individuals through iteration of social interaction [8]. Our simulation results also support his claim.

Tomasello furthermore advocates that the following is important for acquisition of habitual use of linguistic symbol [8]. An individual 1) understands that the others are individuals with some intents, 2) participates in a joint attention situation, 3) comprehends the other's intent in such situation, and 4) can use a symbol that others used toward the individual. Although our reinforcement learning agents did not explicitly have such functions, the simple two-person setup of the game probably made it unnecessary to use attentive mechanisms.

6 Conclusion

We have proposed an intrusion game (IG) for investigating the emergence of protocommunication from a world where a teacher of communication or even any dedicated mechanisms for communication do not exist.

Using computer simulations of IG, we have shown that agents who can turn on/off their lights as redundant actions became able to spontaneously acquire meanings of light signals and cooperate with the other agent. We have also found that agents with working memories can differentiate their roles as a sender and a receiver. Further, our simulation demonstrated that a cooperation without communication can emerge from interaction between the agents having both signaling and memory capabilities.

Our simulation results suggest that repeated interaction between individuals with a reinforcement learning function can play an important role in establishing protocommunication, even if individuals do not have dedicated mechanism for communication.

References

[1] T. Hashimoto, *Knowledge Science* (in Japanese), K. Sugiyama, *et al.* (Eds.), Kinokuniya shoten, pp.126–131, 2002.

[2] A. Cangelosi, D. Parisi, "The emergence of a language in an evolving population of neural networks," *Connection Science*, 10(2), pp.83–97, 1998.

[3] D. Marocco, S. Nolfi, "Emergence of communication in embodied agents: co-adapting communicative and non-communicative behaviours," A. Cangelosi, *et al.* (Eds.), *Modeling language, cognition and action: Proceedings of the 9th Neural Computation and Psychology Workshop*, 2004.

[4] L. Steels, P. Vogt, "Grounding adaptive language games in robotic agents," C. Husbands, I. Harvey, (Eds.), *Proceedings of the Fourth European Conference on Artificial Life*, MIT Press, 1997.

[5] S. Tensho, *et al.*, "Gradual emergence of communication in a multi-agent environment" (in Japanese), *IPSJ SIG-ICS*, 139, pp.73–78, 2005.

[6] R. S. Sutton, A. G. Barto, *Reinforcement Learning*, MIT Press, 1998.

[7] T. R. Halliday, P. J. B. Slater, (Eds.), *Animal Behavior*, Blackwell Scientific Publications, 1983.

[8] M. Tomasello, *The Cultural Origins of Human Cognition*, Harvard Univ. Press, 1999.