Artif Life Robotics (2008) 12:70–74
DOI 10.1007/s10015-007-0444-x
© ISAROB 2008

ORIGINAL ARTICLE

Takashi Sato · Eiji Uchibe · Kenji Doya

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

Received and accepted: May 14, 2007

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 shows that agents with lights and light sensors can learn turn-taking behavior for avoiding collisions using visual communication. Further analysis reveals a variety in the mapping of messages to signals. In some cases, the differentiation of roles into a sender and a receiver was observed. The result confirmed that protocommunication can emerge through interaction between agents having generic reinforcement learning capability.

Key words Reinforcement learning · Intrusion game · Emergence of protocommunication · Role differentiation

1 Introduction

What is the origin of the prototypes of symbolic communication such as linguistic interaction or protocommunication? Then how, and based on what capacity of individuals, did protocommunication emerge? These questions have been discussed in various fields for a long time, but unlike other questions in archaeology, these questions are hard to answer as no evidence of such communication emergence existed until the recent emergence of written languages. Thus, we address these questions by "understanding by construction" using mathematical modeling and computer simulations.[1]

T. Sato (✉) · E. Uchibe · K. Doya
Media Information Engineering, Okinawa National College of Technology, 905 Aza-Henoko, Nago 905-2192, Japan
Tel. +81-980-55-4179
e-mail: stakashi@okinawa-ct.ac.jp

This work was presented in part at the 12th International Symposium on Artificial Life and Robotics, Oita, Japan, January 25–27, 2007

The studies on the 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, the major limitation in most previous studies is that they assumed the preexistence of a basic framework for communication, such as signals and meanings, or a speaker and a listener, and simply verified the evolution or learning of the mappings between signals and meanings by considering the success of communication itself as the objective function. Therefore, it is difficult for these studies to describe how communication emerged from a world where concepts such as signals, words, and speaking did not exist.

The purpose of this study is to discover if communication can emerge between individuals who have basic behavior–learning functions, but do not have any dedicated mechanisms or an absolute need for communication. Specifically, we perform 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 at the same time avoiding collisions. We evaluate what action, sensation, and memory capacities are necessary for the learning of cooperative behavior, and when learned, what meanings agents assign to their excess degrees of action and sensation. Further, we investigate the developmental process of cooperative behavior by communication, and the cases of role differentiation into a speaker and a listener.

2 Intrusion game

We consider an intrusion game (IG) which is a simplification of situations such as a turf war between foraging animals. Two players can move back and forth in a one-dimensional space with four slots. The players are bounded by walls on the "west" and "east" ends of the track, and cannot jump over a slot or stay together in the same slot. Figure 1 depicts six possible sets of positions that the players can assume. We denote each of the six position patterns with the numbers 0–5. The "west" player can get a reward

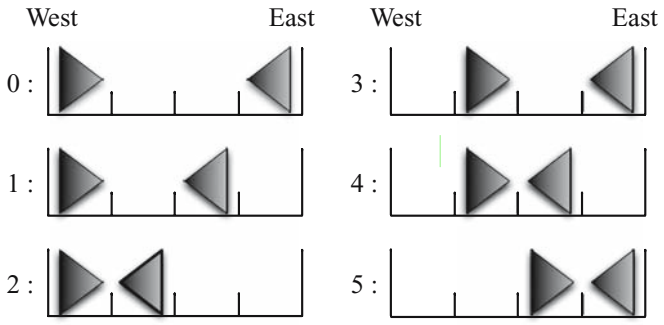

Fig. 1. Six possible position patterns (denoted 0–5) of the players in the intrusion game

by entering the east half of the track (i.e., position pattern 5), and the "east" player by entering the west half (i.e., position pattern 2), without a collision. A penalty (negative reward) is given when a player collides with a wall or another player.

The crucial problem in this game is how the players resolve the conflict at position pattern 4. If the players act selfishly, i.e., each tries to maximize its own reward, then both would take an action to move forward; however, this will cause a collision resulting in negative rewards for both players.

3 Reinforcement learning agents

In order to test whether agents with a general action-learning capability can also learn to communicate, we adopt reinforcement learning agents,[6] which can learn various behaviors based on rewards and penalties. We use the Q-learning method,[6] which is standard for discrete tasks such as 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 the learning rate ($0 < \alpha \leq 1$), γ is the discount rate ($0 \leq \gamma \leq 1$), and r_t is the reward given after action a_t is taken at state s_t. We use the ε-greedy policy in which an action is randomly selected with probability ε, or an action that maximizes the Q-value for a given state is selected.

In order to investigate how the sensory, action, and memory capabilities of the agents affect the learned behaviors, we tested four types of agents. A null or N-type agent simply has two movement actions (backward and forward) and can sense the position pattern (0–5) of the two agents. A light-capable or L-type agent can turn its lights on or off, and also has a light sensor to see if the other agent's light is on or off. A memory-based or M-type agent maintains a 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 60 simulation runs for each of the four types of agent with the following setups: positive reward +1 for a successful intrusion, negative reward −1 each for a collision with a wall or with the other agent, $\varepsilon = 0.01$, $\alpha = 0.01$, $\gamma = 0.9$, maximum episode duration = 10, and 1 episode = 10000 steps.

4.1 Agent behavior

First, we present examples of the typical behavioral patterns (Fig. 2). Figure 2a shows noncooperative dominance by one agent. In this behavioral pattern, one agent can earn a positive reward every two steps. The other can get no reward, but can receive a negative reward if it alters its behavior. Figure 2b presents 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, while the other can obtain only one. Figure 2c depicts suboptimal cooperation leading to one reward every six steps. Figure 2d shows optimal cooperation in which both agents earn a reward every four steps.

We analyze the occurrence frequency of the four typical behavioral patterns for the four types of agent. As can be seen in Fig. 3, the agents without lights (N- and M-type) can only learn noncooperative dominance. In contrast, the agents with lights (L- and LM-type) can demonstrate various types of cooperation. Further, the LM-type agents can achieve optimal cooperation more frequently than the L-type agents, which lack any memory of their previous actions.

4.2 Developmental process

Next, we examine how the behaviors change by learning, before converging to one of the four typical patterns. Figure 4 shows the developmental history of four LM-type agents who have acquired one of the four typical behavioral patterns at the final episode. We recorded the Q-values of the agents every 1000 steps, and let the agents play the IG (with $\varepsilon = \alpha = 0$) starting from all possible initial states.

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

4.3 Variety in signaling

We observed the emergence of various types of communication, even for the same behavioral pattern. Figure 5

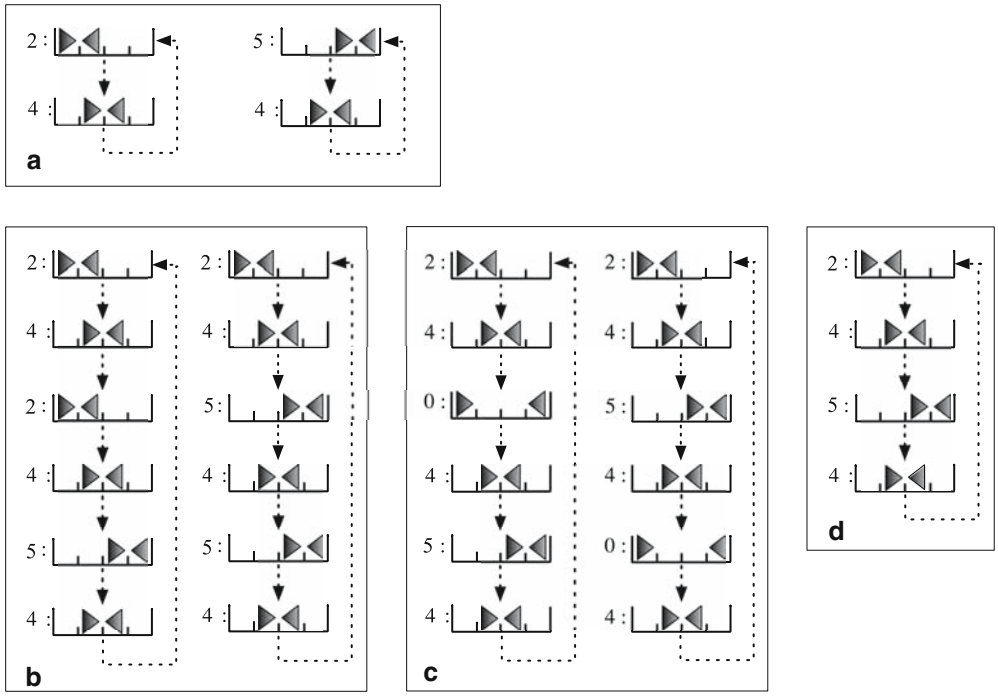

Fig. 2. Four examples of typical behavioral patterns. **a** Noncooperative dominance by one agent. **b** Asymmetric cooperation. **c** Suboptimal cooperation. **d** Optimal cooperation

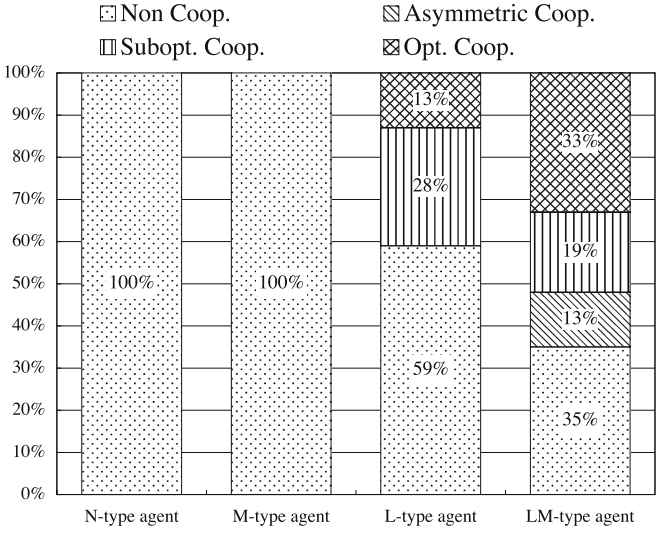

Fig. 3. Occurrence frequency of four typical behavioral patterns for four types of agent

exemplifies four typical types of communication that realize optimal cooperation.

Figure 5a shows an example of symmetric signaling in which the agents can resolve the conflict at position pattern 4 by alternately turning on their lights while stepping forward. This means that the agents can convey their next action as a message via their lights. Figure 5b shows an example of asymmetric signaling in which one agent turns its light on to step forward, while the other agent turns its light on to step backward. These could be seen in both the L- and LM-type agents.

The examples shown in Figs. 5c and d could be seen only in the case of the LM-type agents. Figure 5c depicts one-way communication between the sender (the east agent) and the receiver (the west agent) after 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 signals. The east agent, who cannot rely on the light signals of the west agent, determines its actions based on the memory of its own previous actions in order to solve the conflict at position pattern 4. Figure 5d is a case of optimal cooperation without communication. As can be seen, the east agent always turns on its light when entering position pattern 4 from positions 2 and 5. Therefore, the west agent cannot solve the conflict at position pattern 4 by using the east agent's light signal. The behavior of both agents is based only on the memory of their past actions.

5 Discussion

Our simulation results showed that a variety of communication patterns can emerge through interactions between agents exploring the roles of their redundant actions by a generic reinforcement learning process.

Animal communication is defined as the signal transmission, that senders can profit by the reactions of receivers on average.[7] Our simulation results confirm that communication (defined in this manner) can emerge from repeated interactions between reinforcement learning agents with sufficient physical capabilities. Tomasello[8] claims that communication signals can be generated between two individuals who form each other's behaviors through repeated social interaction. Our simulation results also support this claim.

Tomasello[8] further advocates that the following attentive mechanisms are important for the acquisition of the habit-

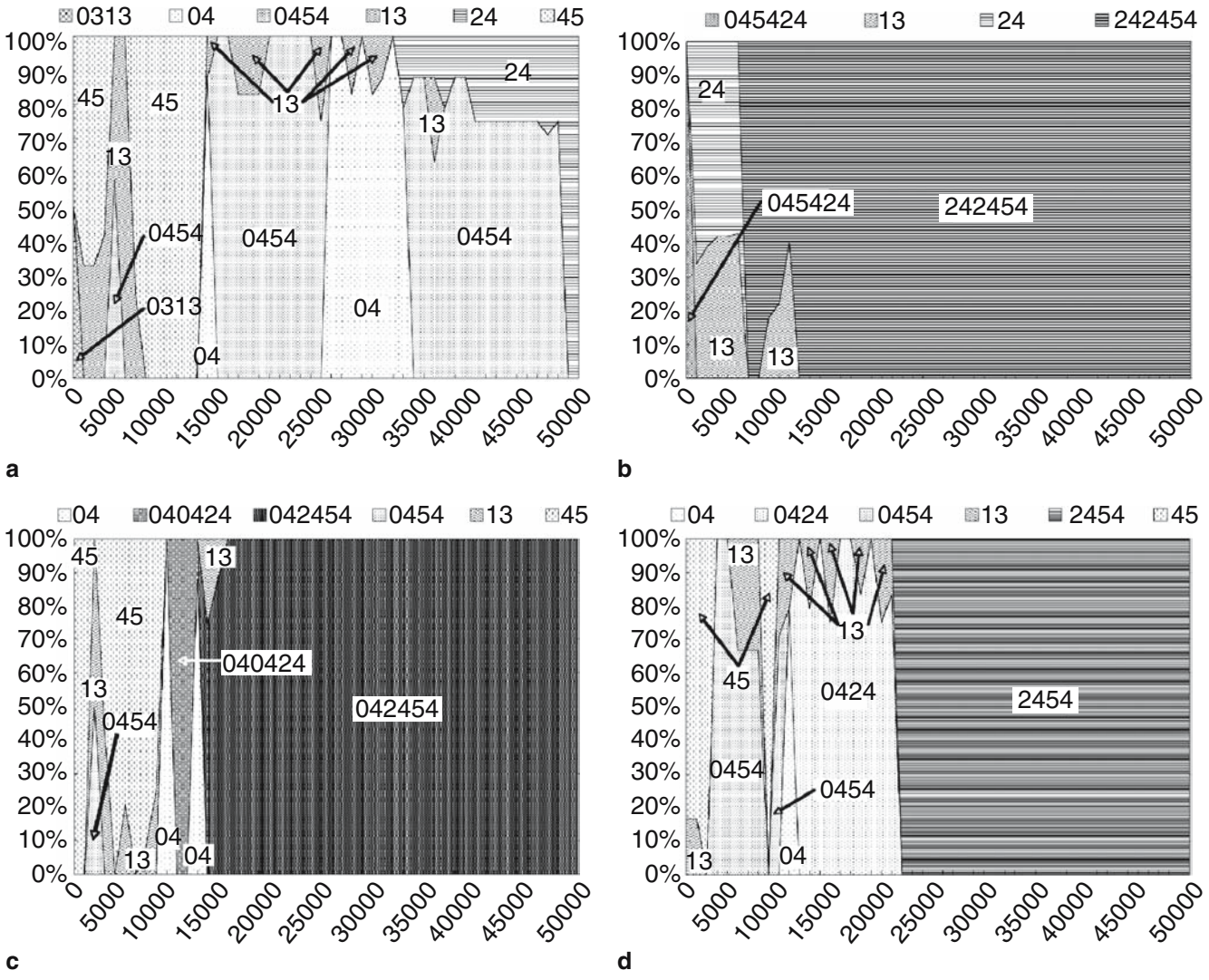

Fig. 4. Examples of the developmental history of behavioral patterns in LM-type agents. The *x*-axis and the *y*-axis of each figure denote the steps and the occurrence frequency of the behavioral patterns, respectively. Each numerical string represents a cyclic position pattern that has converged from 384 different initial states (6 position patterns × 4 light states × 16 memory states). **a** Noncooperative dominance. **b** Asymmetric cooperation. **c** Suboptional cooperation. **d** Optional cooperation

ual use of linguistic symbols. The individual must (1) understand that the others are individuals with some intent, (2) participate in a joint attention situation, (3) comprehend the other's intent in such a situation, and (4) use symbols that others use toward it based on imitative learning. 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) to investigate 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 the IG, we have shown that agents who can turn their lights on/off as redundant actions became able to spontaneously acquire meanings for the light signals and cooperate with each other. We have also found that agents with working memories can differentiate their roles as a sender or a receiver. Further, our simulation has demonstrated that cooperation without any communication can emerge from interactions between agents having both signaling and memory capabilities.

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

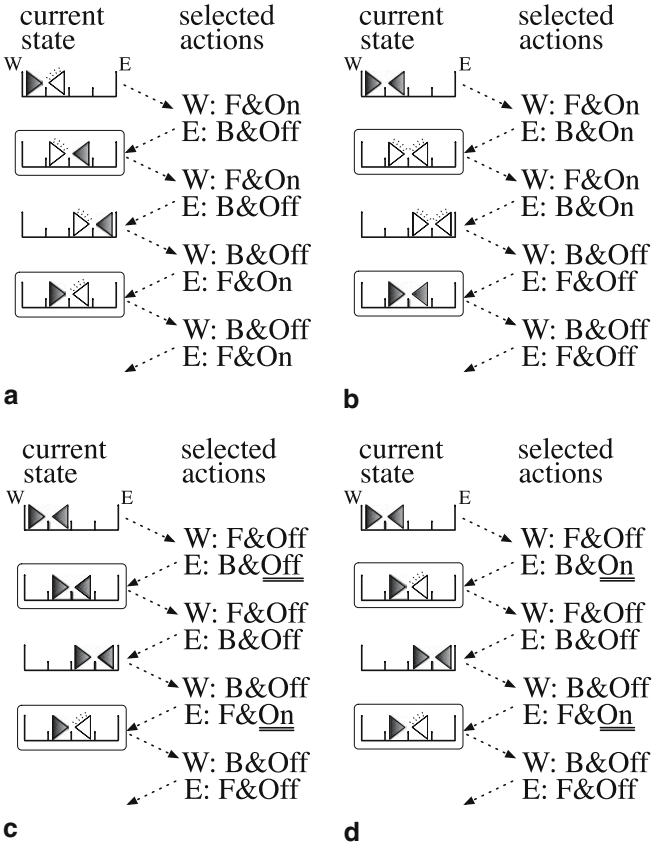

Fig. 5. Typical examples of emerged communication. **a** Symmetric signaling. **b** Asymmetric signaling. **c** One-way communication between a sender and a receiver after role differentiation. **d** Cooperation without communication. *W*, west agent; *E*, east agent; *F*, go forward; *B* go backward; *Off*, light off; *On*, light on

References

1. Kaneko K, Tsuda I (2000) Complex systems: chaos and beyond. A constructive approach with applications in life sciences. Springer, Berlin
2. Cangelosi A, Parisi D (1998) The emergence of a language in an evolving population of neural networks. Connection Sci 10(2): 83–97
3. Marocco D, Nolfi S (2004) Emergence of communication in embodied agents: co-adapting communicative and non-communicative behaviours. In: Cangelosi A et al. (eds) Proceedings of the 9th Neural Computation and Psychology Workshop, Plymouth, Sep. 8–10, 2004, World Scientific, Singapore
4. Steels L, Vogt P (1997) Grounding adaptive language games in robotic agents. In: Husbands C, Harvey I (eds) Proceedings of the 4th European Conference on Artificial Life (ECAL '97), MIT Press, Cambridge, pp 474–482
5. Tensho S, Maekawa S, Yoshimoto J, et al. (2005) Gradual emergence of communication in a multi-agent environment. Proceedings of the 10th International Symposium on Artificial Life and Robotics (AROB 10), Beppu, Oita, Japan, Feb. 4–6, 2005, pp 551–554
6. Sutton RS, Barto AG (1998) Reinforcement learning. MIT Press, Cambridge
7. Halliday TR, Slater PJB (1983) Animal behavior. Blackwell Scientific, Oxford
8. Tomasello M (1999) The cultural origins of human cognition. Harvard University Press, Cambridge