# Emergence of Cooperation in Complex Agent Networks Based on Expectation of Cooperation

## (Extended Abstract)

Ryosuke Shibusawa, Tomoaki Otsuka, and Toshiharu Sugawara*

Deptartment of Computer Science and Communications Engineering
Waseda University, Tokyo 169-8555, Japan
{r.shibusawa, t.otsuka}@isl.cs.waseda.ac.jp, sugawara@waseda.jp

## ABSTRACT

This paper proposes a behavioral strategy called expectation of cooperation strategy with which cooperation in the prisoner's dilemma game emerges in agent networks by incorporating Q-learning. The proposed strategy is simple and easy to implement but nevertheless can evolve and maintain cooperation in all agent networks under certain conditions. We conducted a number of experiments to clarify these conditions, and the results indicate that cooperation emerged.

## Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems*

## General Terms

Experimentation

## Keywords

Norm, Complex networks, prisoner's dilemma

## 1. INTRODUCTION

Humans are basically self-interested and individually rational but often pursue social benefits on the basis of, for example, altruism and indirect reciprocity. This kind of behavior is selected to avoid/resolve social dilemmas such as conflicts between (groups of) humans and the existence of free riders. Agents' decisions made only from the self-interested and local viewpoints often result in non-cooperative behavior resulting in conflicts that may result in loss for both sides. To avoid such conflicts, agents should take into account not only local benefits but also social benefits; such socially cooperative behavior may be the non-best choice from a short-term and local viewpoint but will bring better results eventually in the long term.

A number of studies have been conducted, and socially cooperative behavior as a norm has emerged in which Nash equilibria are not appropriate from social viewpoint [2, 3, 1]. However, these studies did not succeed in the emergence of cooperation in social dilemma in some agent networks or it is not obvious to implement in the actual systems.

The contribution of this paper is the proposal of the simple behavioral strategy called *expectation of cooperation* with which (the norm of) cooperation in the prisoner's dilemma (PD) games emerges under a certain condition. We then experimentally show that the expectation of cooperation strategy fosters cooperation in the whole networks, although it is simple and easy to implement.

*This work is supported by KAKENHI (25280087).

**Appears in:** *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*


## 2. PROPOSED METHOD

### 2.1 PD Games with Ensemble Strategy

For a given agent network $G = (A, E)$, we assume that all pairs of agents in $E$ play the PD game once in a single round in random order. Agent $i$ decides its strategy by the *ensemble strategy decision with Q-value* or simply the *ensemble decision* with $\varepsilon$-greedy method for all neighbors. More specifically, $i$ has the Q-value, $Q^i(s, j)$, for the strategy $s \in S_{tr}$ and neighbor agent $j \in N_i$ at the $t$-th round, and it is updated by

$$Q^i_t(s, j) = (1 - \alpha_i) \cdot Q^i_{t-1}(s, j) + \alpha_i \cdot r^i_j(t - 1),$$

where $r^i_j(t-1)$ is the payoff received after the latest PD game with $j$ at round $t-1$, and $\alpha_i$ is the parameter for the learning rate of agent $i$.

At the beginning of the $t$-th round, agent $i$ identifies its strategy, $s^i(t)$, so that the following condition is satisfied.

$$s^i(t) = \arg\max_{s \in S_{tr}} p^i(s), \quad (1)$$

where $p^i(s)$ is the preference function for ensemble decision. In our experiments, we define it as a plurality, i.e.,

$$p(s) = \sum_{j \in N_i} \delta(s, s^i_j(t-1)) \quad (2)$$

$$s^i_j(t-1) = \arg\max_{s \in S_{tr}} Q^i_{t-1}(s, j), \quad (3)$$

where $\delta(s, s') = 1$ if $s = s'$; otherwise it is zero. Then, $i$ selects strategy $s^i(t)$ with probability $1 - \varepsilon$ and selects it randomly from $S_{tr}$ with probability $\varepsilon$. We call strategy $s^i(t)$ as the *preferred strategy* of $i$ at the $t$-th round.

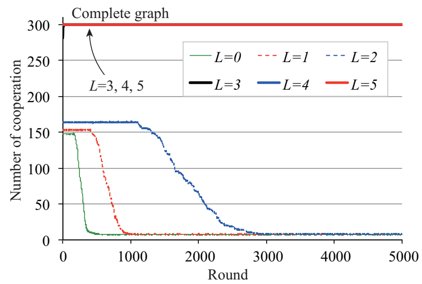


**Figure 1: Ratios of cooperation per round.**

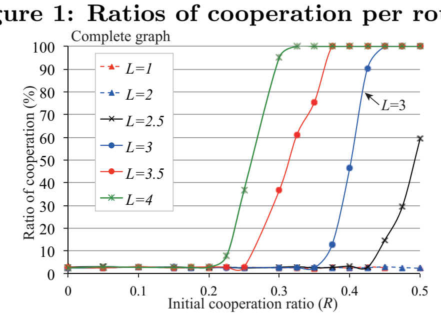


**Figure 2: Ratios of cooperation.**

## 2.2 Expectation of Cooperation Strategy

Agents with the proposed behavioral strategy, called the *expectation of cooperation strategy*, are assumed to know that social cooperative behavior is better than defection, so they expect (hope for) it when interacting with their neighbors, but they also know that it is fruitless if other local agents do not cooperate. Furthermore, agents can observe others' strategies only when they interact. We introduce the positive integer, $L$, which represents *cooperation persistence*. Agent $i$ also has the parameter $l_i (\geq 0)$ to express the remaining term for persisting cooperation, i.e., when $l_i = 0$, $i$ selects the preferred strategy, but while $l_i > 0$, $i$ causes cooperation persistence in the hope of cooperation in the local environment. The initial value of $l_i$ is set to zero.

If agents $i$ and $j$ mutually select cooperation, $(C, C)$, occasionally, they will individually set $l_i$ and $l_j$ to $L$ and enter the term for persisting cooperation in the hope of (local) emergence of cooperation. Thus, $i$ and $j$ persist cooperation in at least the next $L$ PD games (not rounds) with their neighbors, and $l_i$ is decremented by one after each game. However, if $i$ encounters joint $(C, C)$ again, $l_i$ is set to $L$ and the term of persisting cooperation is prolonged.

It is obvious that if the cooperation persistence, $L$, is large, agents continue to select $C$, and if $L = 0$, it is the PD games with normal Q-learning, so all agents tend to select $D$, which is the Nash equilibrium. We are interested in the smallest value of $L$ that leads to (almost) total cooperation.

# 3. EXPERIMENTAL RESULT

We evaluated our strategy using the agent network of compete graph. The payoff matrix for PD game is defined as

| | $C$ | $D$ |
|---|---|---|
| $C$ | 3, 3 | 0, 5 |
| $D$ | 5, 0 | 1, 1 |

The learning rate $\alpha$ ($= \alpha_i$ for $\forall i \in A$) is set to 0.1, and $\varepsilon$ for the $\varepsilon$-greedy strategy is set to 0.05. The number of agents $|A|$ is three hundred. The data described below is the mean values of 20 independent trials.

Figure 1 plots the ratios of agents that select cooperation, $C$, per round. Note that the initial values of $Q_0^i(s, j)$ for $\forall i \in A$ and $\forall j \in N_i$ is set to 0. We can see that when $L \leq 2$, cooperation could not emerge in the society of agents with the expectation of cooperation strategy but did emerge (thus the preferred strategy, $s^i(t)$, converged to $C$ for $\forall i \in A$) when $L \geq 3$. Note that we also conducted the same experiment for the agents with only expectation of cooperation strategy without the ensemble strategy decision with Q-values, so agent $i$ selects $D$ with probability $1 - \varepsilon$ and $C$ or $D$ randomly with probability $\varepsilon$ when $l_i = 0$. In this case, no cooperation emerged even when $L = 5$.

The ratios of cooperation seemed to depend on the ratios of cooperation in the first round (initial strategies) due to the characteristics of expectation of cooperation behavioral strategy since the joint action $(C, C)$ by a pair of agents forced them to cooperate individually for the next $L$ rounds. In the experiment described above, $Q(C) = Q(D) = 0$, so fifty percent of the agents preferred cooperation ($s^i(0) = C$), and the other agents preferred defection ($s^i(0) = D$) on average. Thus, we set the Q-values, $Q(C) = Q(D) = 0$, like in the previous experiment, but we assumed that agents selected $C$ as the initial preferred strategy, $s^i(0)$, with probability $R$ ($0 \leq R \leq 0.5$, so the other agents selected $D$ with probability $1 - R$).

Figure 2 plotted the ratios of cooperation at the 5000th round when $R$ was varied. The ratios of cooperation sharply increase to 100% with increasing initial cooperation rate, $R$. For example, the curve for $L = 3$ indicated that all agents selected $C$ after all when $R = 0.45$, but approximately half of the agents and none of the agents selected $C$ when $R = 0.40$ and $R = 0.35$, respectively.

When the initial cooperation ratio was small, the experiment started from the situation in which many agents preferred $D$. Furthermore, since agents were tightly connected to each other in the complete graph, relatively larger cooperation persistence, $L$, was necessary to overturn the situation where $D$ was the majority strategy, but the overturn was immediate if $R$ reached a sufficiently large number.

# 4. CONCLUSION

We proposed the strategy called expectation of cooperation in which agents continue to cooperate for $L > 0$ times after mutual cooperation. We also conducted other types of networks, such BA and CNN, and if agents select $C$ or $D$ randomly, $L = 3$ seems the minimal number for this strategy from our experiments.

# REFERENCES

[1] N. Masuda. Evolution of cooperation driven by zealots. *Nature Scientific Reports*, 2(646), 2012.

[2] K. Moriyama, S. Kurihara, and M. Numao. Evolving subjective utilities: Prisoner's dilemma game examples. In *Proc. of AAMAS 2011 - Vol. 1*, pages 233–240, 2011.

[3] D.-M. Shi, H.-X. Yang, M.-B. Hu, W.-B. Du, B.-H. Wang, and X.-B. Cao. Preferential Selection Promotes Cooperation in a Spatial Public Goods Game. *Physica A: Statistical Mechanics and its Applications*, 388(21):4646 – 4650, 2009.