Reinforcement Learning Explains Conditional Cooperation in Repeated Social Dilemma Games

Takahiro EZAKI[1,2], Yutaka HORITA[3], Masanori TAKEZAWA[4,5] and Naoki MASUDA*[6]

[1]National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. [2]JST, ERATO, Kawarabayashi Large Graph Project, Global Research Center for Big Data Mathematics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. [3]Department of Psychology, Teikyo University, 359 Otsuka, Hachioji, Tokyo 192-0395, Japan. [4]Department of Behavioral Science, Hokkaido University, N10W7, Kita-ku, Sapporo 060-0810, Japan. [5]Center for Experimental Research in Social Sciences, Hokkaido University, N10W7, Kita-ku, Sapporo 060-0810, Japan. [6]Department of Engineering Mathematics, University of Bristol, Merchant Venturers Building, Woodland Road, Clifton, Bristol BS8 1UB, United Kingdom.

E-mail: *naoki.masuda@bristol.ac.uk

Keyword: conditional cooperation, direct reciprocity, reinforcement learning, prisoner's dilemma, public goods game

Behavioral experiments using repeated social dilemma games in groups and networks have revealed that players' decisions are strongly correlated to the fraction of cooperative partners in the previous round of the repeated game [1, 2]. As the fraction of cooperative partners increases, on average, players tend to cooperate with a higher probability (squares in Fig. 1), which is referred to as the conditional cooperation. When the focal player has cooperated in the previous round, this pattern is remarkable (circles in Fig. 1), while the probability of cooperation is less affected or even decreased by an increase in the fraction of cooperative neighbors when the player has defected in the previous round (triangles in Fig. 1). As a more detailed description, these behavioral patterns are referred to as the moody conditional cooperation. The origin of conditional cooperation and its moody variant largely remains unclear. We provide a proximate explanation by numerical simulations. We found that players adopting a variant of the so-called Bush-Mosteller reinforcement learning rule [3] show the targeted behavior as shown in Fig. 1. In the model, players respond only to the payoff they gained in the previous round and have no access to information about the actions of other players. Thus they cannot explicitly use conditional cooperation rules. We found that the reinforcement learners that showed moody and non-moody conditional cooperation obeyed a behavioral pattern similar to the GRIM strategy, which is a well-known strategy in the repeated prisoner's dilemma game. A reinforcement-learning variant of the GRIM strategy seems to better explain the past experimental results than the Pavlov strategy, an established strong competitor in the repeated prisoner's dilemma game.

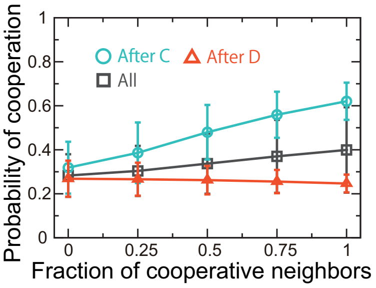

Figure 1: Conditional cooperation (squares) and moody conditional cooperation (circles and triangles) in the repeated prisoner's dilemma game. Probability of cooperation is plotted against the fraction of cooperative neighbors in the previous round. The squares represent the results not conditioned on the previous action of the focal player. The circles and triangles represent the probability to cooperate when the player has cooperated and defected in the previous round, respectively. These results were obtained by a variant of the Bush-Mosteller model.

References

[1] U. Fischbacher, S. Gächter, E. Fehr, "Are people conditionally cooperative? Evidence from a public goods experiment", Econ. Lett. vol. 71, 397 (2001).

[2] J. Grujić, C. Fosco, L. Araujo, J.A. Cuesta, A. Sánchez, "Social experiments in the mesoscale: humans playing a spatial prisoner's dilemma", PLOS ONE vol. 5, e13749 (2010).

[3] R.R. Bush, F. Mosteller, *Stochastic Models for Learning*, Wiley, New York (1955).