Theory of Cooperation: Exploring Human and Machine Capabilities

Naoto Iwahashi — Okayama Prefectural Univ
Hiroyuki Okada — Tamagawa Univ
Kotaro Funakoshi — Tokyo Tech

Abstract

Cooperation is one of the most fundamental human characteristics, and several interdisciplinary studies have been conducted to understand this aspect. However, despite the numerous studies conducted on the subject, a comprehensive theory regarding cooperation remains elusive. In this study, the theoretical and practical aspects of human–human, human–machine, and machine–machine cooperation are explored. Cooperation is theoretically analyzed from behavioral, mathematical, and cognitive viewpoints. The theoretical principles were applied to certain unsolved problems. Furthermore, it is demonstrated that such cooperative machines can be used to investigate the human capabilities of cooperation. The developed versatile interactive software *RoCoCo*, which can be used as a research tool, is presented. Finally, the possibilities and prospects of this research framework, which is termed the *theory of cooperation*, are discussed.

Keywords: cooperation, theory, practice, human, machine, RoCoCo

1 Introduction

Cooperation is a fundamental and crucial human characteristic, regarding which numerous interdisciplinary studies (including philosophy, science, and technology) have been conducted. Despite the recent rapid research progress, a comprehensive theory regarding cooperation has not been established yet. In this study, we introduce a framework, termed as the *theory of cooperation*, which we proposed to explore human–human, human–machine, and machine–machine cooperation, both via theoretical research and development of practical machines.

The *theory of cooperation* aims to answer the following questions:

- What is cooperation?
- How does cooperation work?
- What is human cooperation ability?
- How do we apply cooperation to artificial intelligence systems?

2 Principles of cooperation

We analyze cooperation from behavioral, mathematical, and cognitive viewpoints to explore the principles that establish cooperation.

2.1 Behavioral

From a behavioral viewpoint, cooperation generally comprises the following three interdependent and hierarchical behavioral functions:

Action This function enables each agent to act according to the given dynamics and constraints.

Interaction This function enables agents to mutually predict each other's behaviors and act under a given role assignment.

Role coordination This function enables agents to mutually find and select their own role assignments online.

Among these three functions, *role coordination* is particularly important and has been investigated in several studies. However, this function is not yet completely understood with regard to both cognitive and mathematical aspects.

2.2 Mathematical

We propose hierarchical equilibrium dynamics (HED) as a mathematical formulation of general cooperation. It is based on control, game theory, and physics, and characterized by the structure of dynamics and equilibrium of coupled dynamics.

The structure comprises the following three-layered dynamics according to the behavioral hierarchies (Fig. 1):

Action and task dynamics Modeling of the actions and tasks executed by the agents.

Interaction equilibrium dynamics Modeling of the interactions among agents, which are described via the occurrence of an equilibrium in the coupling of dynamical systems.

Role equilibrium dynamics Modeling of role coordination, which controls the action equilibrium dynamics.

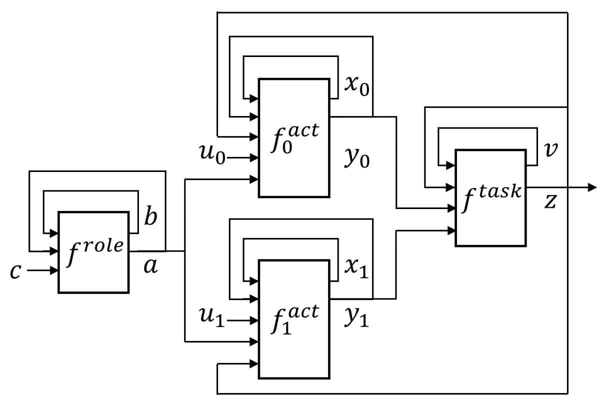

Figure 1: Structure of dynamics that establish cooperation

The dynamics of the role coordination for two agents, f^{role}, and the dynamics of the action of the ith agent, f_i^{act}, can be expressed as follows:

$$a_{t+1}, b_{t+1} = f^{role}(a_t, b_t, c_t)$$
$$y_{i,t+1}, x_{i,t+1} = f_i^{act}(y_{i,t}, x_{i,t}, u_{i,t}, a_t, z_t)$$

The task dynamics, f^{task}, can be expressed as follows:

$$z_{t+1}, v_{t+1} = f^{task}(z_t, v_t, y_{0,t}, y_{1,t})$$

For optimization, the objective for the control of the ith agent is expressed as follows:

$$J_i^{overall} = J^{role}(A, B, C) + J_i^{act}(Y_i, X_i, U_i) + J^{task}(Z)$$

where J^{role}, J_i^{act}, and J^{task} denote the objectives for the control of the dynamics f^{role}, f_i^{act}, and f^{task}, respectively.

HED specializes or generalizes a wide range of formulas for synchronization, such as non-cooperative game [1], differential game [2], decentralized control [3], pulse-coupled biological oscillators [4], chaotic synchronization [5], coupling of recurrent neural networks [6, 7], coupling of statistical dynamical models [8], and recursive reasoning [9].

Previous studies on cooperative dynamics have addressed the action and interaction functions but not the role coordination function. In particular, studies on the interaction function focused on convergence to an equilibrium point. However, owing to the existence of multiple or infinite equilibrium points in typical problems, emphasis should be placed on the selection of an appropriate equilibrium point among the available points; this selection is defined as role coordination.

2.3 Cognitive

Similar to the equilibrium in game theory that is based on mutual belief, the equilibrium in HED is also based on mutual belief. Moreover, because the equilibrium in HED is hierarchical, a hierarchical structure of mutual beliefs is required. To meet this requirement, we

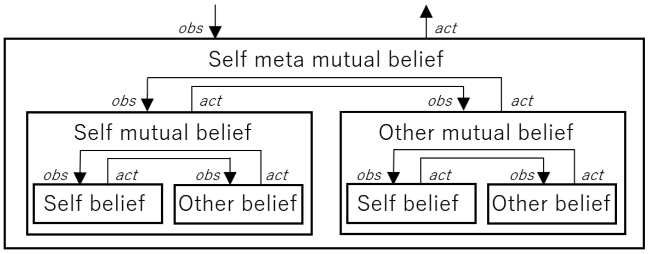

Figure 2: Belief configuration of an agent for cooperation

propose a mutual belief hierarchy with three layers, each of which is represented by the corresponding dynamics as follows:

Belief The dynamics used by an agent to make decisions.

Mutual belief The dynamics of the coupling between the belief of an agent and those of others; not mutual beliefs themselves.

Meta mutual belief The dynamics of the coupling between the mutual beliefs of an agent and those of others; not meta mutual beliefs themselves.

The belief configuration of an agent with this belief hierarchy system is depicted in Fig. 2.

The proposed hierarchical mutual belief system differs from the traditional mutual belief system [10] in two ways. First, the coordinating agents' different mutual beliefs are emphasized. As our mutual beliefs never become same, the most important problem to be solved should be the achievement of cooperation even under different mutual beliefs. However, traditional theories of joint activities do not solve this problem. The cognition of an agent should allows the agent to coordinate her/his mutual belief with the mutual beliefs of others. Meta mutual belief is set at the top of the hierarchy of the proposed system and provides the important ability to coordinate roles. Second, mental self-recursive reflection is represented by the equilibrium of dynamics. The traditional mutual belief system is represented by the metaphor of self-recursive reflection by mirrors. However, it seems impossible to construct the mechanism of self-recursive reflection in the brain from the viewpoints of information and control theories, and physics. Therefore, we use the concept of equilibrium between dynamics instead, which seems feasible in the brain.

3 Machine study

We applied the above-mentioned principles to the following unsolved problems: autonomous driving [11], human–robot physical interaction, multi-modal language acquisition [12, 13], generation [14], and understanding[15], and robot-directed utterance detection [16]. From these applications, the validity of the principles was confirmed.

	Cooperation	Control
Human	feasible	infeasible
s-r machine	infeasible	feasible
Cooperative machine	feasible	feasible

Table 1: Comparison of the characteristics of three different experimental frameworks in which subjects interact with humans, s-r (stimulus-responsive) machines, and cooperative machines, respectively.

4 Human study with cooperative machines

The above-mentioned cooperative machines can be used to investigate the human capabilities of cooperation. In conventional experiments, human subjects interact with human or stimulus-responsive (s-r) machines. However, there are some limitations to achieving coordinated operation and reproducible conditional control of the experiments (Table 1). By using cooperative machines, we can reduce such limitations, and explore human capabilities more accurately and deeply.

5 Research tool

We developed the interactive software, RoCoCo, as a research tool for exploring human and machine capabilities. RoCoCo has the following features:

- Real time interactive
- Versatile and expandable
- Multi-purpose (human and machine studies)
- Flexible configuration (number of agents, human-machine, human-human, and machine-machine)
- Flexible task setting (autonomous driving and physical interaction)
- Various quantitative evaluations

The RoCoCo is now expanding in its versatality and flexibility.

6 Conclusions

On the basis of the fundamental analysis of cooperation, cognitive and mathematical principles were proposed. The validity of the principles was verified by applying them to certain unsolved tasks. The potential of cooperative machines for exploring the human capabilities of cooperation was discussed. A human study using the interactive software RoCoCo is underway.

References

[1] J. Nash, “Non-cooperative games,” *Annals of mathematics*, pp. 286–295, 1951.

[2] D. Wishart, “Differential games. a mathematical theory with applications to warfare and pursuit, control and optimization,” *Physics Bulletin*, 1966.

[3] N. Sandell, *et al.*, “Survey of decentralized control methods for large scale systems,” *Automatic Control*, 1978.

[4] R. E. Mirollo and S. H. Strogatz, “Synchronization of pulse-coupled biological oscillators,” *SIAM Journal on Applied Mathematics*, 1990.

[5] L. M. Pecora and T. L. Carroll, “Synchronization in chaotic systems,” *Physical review letters*, 1990.

[6] T. Ikegami and M. Taiji, “Imitation and cooperation in coupled dynamical recognizers,” in *European Conference on Artificial Life*. Springer, 1999.

[7] J. Tani, *et al.*, “Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using rnnpb,” *Neural Networks*, 2004.

[8] T. Sasaki, N. Iwahashi, and K. Funakoshi, *et al.*, “Learning and generation of collaborative actions by mdl coupled hmms,” *National Convention of IPSJ*, 2018.

[9] Y. Wen, *et al.*, “Probabilistic recursive reasoning for multi-agent reinforcement learning,” in *Proc. ICLR*, 2019.

[10] H. H. Clark, *Using language*. Cambridge university press, 1996.

[11] N. Iwahashi, “Equilibrium selective role coordination for autonomous driving,” in *Int. Conf. Awareness Science and Technology*, 2019.

[12] ——, “Language acquisition by robots - towards new paradigm of language processing -,” *Journal of JSAI*, 2003.

[13] N. Iwahashi, *et al.*, “Robots that learn to communicate,” in *AAAI Fall Symposium: Dialog with Robots*, 2010.

[14] S. Nakamura, N. Iwahashi, and T. Nagai, “Mutully-adaptive generation of utterances based on estimation of belief shared by human and robots in real world,” *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 21, no. 5, pp. 663–682, 2009.

[15] K. Sugiura and N. Iwahashi, *et al.*, “Object manipulation dialogue by estimating utterance understanding probability in a robot language acquisition framework,” *Journal of the Robotic Society of Japan*, vol. 28, no. 8, pp. 978–988, 2010.

[16] X. Zuo, N. Iwahashi, and K. Funakoshi, *et al.*, “Detecting robot-directed speech by situated understanding in physical interaction,” *Information and Media Technologies*, 2010.