Intelligence Arms Race: Delayed Reward Increases Complexity of Agent Strategies

Hirotaka Osawa
Faculty of Engineering, Information and Systems, University of Tsukuba
1-1-1, Tenno-dai, Tsukuba, Japan
osawa@iit.tsukuba.ac.jp

ABSTRACT
Social brain theory hypothesizes that the human brain becomes larger through evolution mainly because of reading others' intentions in society. Reading opponents' intentions and cooperating with them or outsmarting them results in an intelligence arms race. The author discusses the evolution of such an arms race, represented as finite state automatons, under three distinct payoff schemes and the implications of these results, which suggest that agents increase complexity of their strategies.

Categories and Subject Descriptors
I.6.0 [**Simulation and Modeling**]: General

Keywords: Multi-agent simulation, Genetic programming, Intelligence arms race, Social brain theory

1. INTRODUCTION
The Theory of Mind (ToM) – the brain function for understanding other's intention from the environment - is one of the most complex human skills in cognitive science [1]. The social brain hypothesis in biology states that humans must handle communication with each other for trading benefits in a society [2]. Like the arms race of animal predators and games [3], it is argued that evolutionary pressure from reading an opponent's intention results in an intelligence arms race [4]. If one agent in a group is more intelligent than others, the agent can understand the strategies of others and will cooperate with or outsmart them. As a result, the agent gets more advantages compared with other agents. The process of creating such competitive intelligence is also a challenging theme both in artificial intelligence and multi-agent simulation. If we understand the process of reading other's intentions, artificial systems will be able to understand more intentions and motivations of users.

What kind of trade emerges in society and especially how cooperation is emerging in our society, is discussed in game theory, economics, and artificial life. Axelrod's contest of the iterative prisoner's dilemma game (IPD) and the strength of tit-for-tat (TFT) as a winner is well known [5] and many successor trials have been focused on how a group of agents acquires mutual trust after evolution [6][7][8]. These studies mainly focused on the behavior of society. In other words, they focused on how the entire society forms a cooperative state, and these agent strategies are restrained to a simple level. On the other hand, how each agent acquires ToM during the evolution/simulation process is interesting for artificial intelligence and cognitive science. ToM in animal and human are simulated and implemented [9][10][11]. However, there have been relatively few studies for the intelligence arms race, which is a quick improvement in intelligence through evolution suggested by social brain theory. A key factor is the relationship between evolutionary pressure and our intelligence.

For this research, the author attempted to determine what kind of process will result in an intelligence arms race. The author applied the anti-max prisoner's dilemma game (AMPD), which is an IPD with modified payoff scheme and proposed by Angeline, as a sample task for analyzing above statement [7]. It was conducted to see how mutual and non-mutual trust in trading arises by multi-agent simulation using finite state automatons. However, the automatons used by Angeline were fixed and the factor of intelligence arms race was not evaluated. Osawa et al. evaluated Angeline's AMPD using human-based simulation and found that top ranked agents acquire more automatons [12]. However, this experiment was based on human participants and an increase in intelligence was not automatically derived and not complete for accurate discussion. The author evaluated Angeline's three payoff schemes (IPD, multi-max prisoner's dilemma (MMPD), and AMPD) with artificial evolution in computer simulations of free-scale automatons and evaluated how the number of states and edges of the automatons increased during the game.

The paper is organized as follows. Section 2 describes the simulation model and our hypotheses on the simulations. Section 3 describes the results of simulations with IPD, MMPD, and AMPD and Section 4 discusses our results both through macro-based and micro-based analyses. Section 5 describes the limitations of this study and future work. Section 6 concludes the paper with the results.

2. SIMULATION MODEL
For evaluating the increase in intelligence, the author used evolutional simulations of agents with automatons for strategies. Section 2.1 describes the background of these games by referring to results of game theory, and how our focus differs from them. Section 2.2 gives details of the simulation conditions. Section 2.3 explains how to describe agent strategies by using finite state automatons. Section 2.4 gives the details of the genetic programming (GP) method applied in these simulations. Section 2.5 gives the hypotheses of simulations.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

2.1 Features of IPD, MMPD, and AMPD

The iterative prisoner's dilemma is a typical game in game theory, and it is designed in such a way that the reward is maximized if both players cooperate [5]. A cooperative strategy in IPD is achievable without players having to estimate each other's strategy. This kind of game model is appropriate for simulating ecological behaviors of animals [13]. In a game, each agent has two choices to opponent called "cooperate" show in C and "defect" shown in D.

In IPD, each player has a chance to obtain more rewards for betraying opponents. However, if both players need maximum rewards from the trading, both "cooperate", shown as C-C, is the most appropriate strategy. Consequently, studies on IPD in game theory mainly discuss how to achieve stability created by continuous C-C during the game. Tit-for-tat is one of the simplest and toughest strategies in IPD in which an agent cooperates if the opponent cooperated in the previous round, and defect if the opponent defected in previous round. Axelrod proposed that TFT's quick response and generosity work to maintain cooperation during the game. Later studies revealed that TFT is not a stable strategy [14]. Another strategy called GRIM plays all defect if the opponent defected in the previous rounds. Pavlov is another strategy that changes hand according to the result of the previous round [15]. Pavlov is stronger than TFT if player's hands are informed each other with noise [16].

These strategies can be described by simple rules and are not complex (the author calls them simple based on a previous IPD study [5]). The author believes that this is because the reward is immediately determined in IPD. However, the reward of a trade is sometimes delayed. For example, we can use a credit card for payment if we do not have cash. Human society allows this delayed reward by credit payment because the trader "credits" the opponent's payment in the future. The author estimates that this delayed payment will require more complex ability for each agent. Fisher and Shapiro used iterative arm wrestling for demonstrating delayed trade in a human-based experiment [17]. They demonstrated that if two players play an iterative arm wrestling game and the winner obtains a reward in each match, it is better for both players to fix the game rather than engage in a real fight. They also showed that the key factors in agreeing to fix a game is that each player needs to be intelligent and trust that after if he or she intentionally loses a match, his or her opponent will intentionally lose the next match.

Angeline proposed MMPD and AMPD and took these delayed rewards into account in a game simulation [7]. He modified the IPD payoff table so that it could take into account the mutual trading behavior of Fisher and Shapiro's iterative arm-wrestling game [17]. In MMPD, rewards from cooperative behavior (continuous C-C) are the same as those from mutual defect (continuous C-D and D-C pairs, like one agent plays C, D, C, D,... and another plays D, C, D, C,...). In AMPD, rewards from mutual defect are better than cooperative behavior. Being able to estimate the intentions of other people is important in trading in the real world and requires intelligence. The "intelligence for estimating intention" improves if an agent's reward is not immediately given. The anti-max prisoner's dilemma game can model such trades, and the author argues that AMPD is a suitable game model for verifying the social brain hypothesis [2].

Table 1 is a payoff table of a trading game. The IPD payoff scheme is shown in Eq. 1, the MMPD payoff scheme is shown in Eq. 2, and the AMPD payoff scheme is shown in Eq. 3.

Table 1. Payoff table of Trading Game used in IPD, MMPD, and AMPD.

		Opponent	
		Cooperate	**Defect**
Player	**Cooperate**	$(Pl:c, Op:c)$	$(Pl:b, Op:a)$
	Defect	$(Pl:a, Op:b)$	$(Pl:d, Op:d)$

$$a > c > d > b, \quad a + b < 2c \quad (1)$$

$$a > c > d > b, \quad a + b = 2c \quad (2)$$

$$a > c > d > b, \quad a + b > 2c \quad (3)$$

The author selected a payoff table for IPD as a = 7, b = -3, c = 3, d = -1, that for MMPD as a = 7, b = -3, c = 2, d = -1, and that for AMPD as a = 7, b = -3, c = 1, d = -1 based on previous studies by Angeline and Osawa et al. In the IPD game, C-C hand is Pareto dominate and the average of both players' rewards is maximized as 3 in this case. On the other hand, there is no Pareto dominate hand in AMPD. However, mutual defection in repeated games maximizes the average of both players' rewards. The average is 2 in this case and MMPD.

2.2 Simulation conditions

The author tested 100 IPDs, 100 MMPDs, and 100 AMPDs in the simulations. All agents traded in a round robin fashion during each game. The round robin was repeated 1500 times in one trial and 50 agents battled each other in each round. In each round, one agent battled the rest of the 49 agents (total 2450 battles in each round). The author selected the maximum number of matches in one trade as between 95-104. The maximum number of matches changed randomly for each match to prevent overfitting for fixed matches. The average score of each agent is calculated by the sum of the scores divided by the total number of matches.

2.3 Strategy by finite state automaton

Each agent describes a strategy by using a finite state automaton. Each state in the automaton has numbers representing cooperate and defect of the agent. Even states represent cooperation and odd states represent defect. Several well-known samples are shown in Fig. 1. Each state, or node in Fig. 1, has two edges. These automaton-based notations for strategies are easily applied to the GP method described in the next subsection.

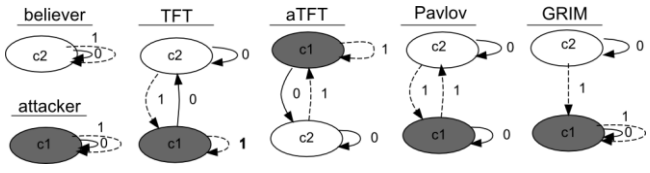

Figure 1. Six sample automatons that describe well known strategies used in game theory. They are shown by automatons with one node and two edges, or two nodes and four edges. Upper node means start point of automaton. The number after 'c' in a circle means state number (cooperative hand is shown as even numbers and defect hand is shown as odd numbers). Number on side of edge shows type of hand (cooperate/C=0 as solid line, defect/D=1 as dashed line).

Each participant describes their strategy using the start state number and several triplets in a simulation program. The transition arrows between nodes are denoted with a set of three numbers (triplet). The first number represents the present state, the second number represents the opponent's hand (0 means cooperate and 1 means defect), and the third state represents the next state (even, odd, or 0 state). Believer and attacker is very simple strategies that are shown in one node and two edges. For example, {{2}, {2,0,2}, {2,1,2}} means a strategy for believer that is cooperative anytime. {{1}, {1,0,1}, {1,1,1}} means attacker that defects opponent anytime. {{2}, {2,0,2}, {2,1,1}, {1,0,2}, {1,1,1}} shows the strategy of TFT, which is common in IPD. {{1}, {2,0,2}, {2,1,1}, {1,0,2}, {1,1,1}} shows the strategy of aTFT. It is an alternative TFT strategy that is almost the same as TFT except it starts from the defect state. {{2}, {2,0,2}, {2,1,1}, {1,0,1}, {1,1,2}} means Pavlov that changes its state from opponent's defect. {{2}, {2,0,2}, {2,1,1}, {1,0,1}, {1,1,1}} means GRIM that plays continuous defects when it is defected.

2.4 Evolution rules on genetic programming

For evaluating the increase size in strategy in three payoff schemes (IPD, MMPD, AMPD), the author applied the GP method to our simulations. In a simulation, 50 agents are living during 1500 rounds/generations. The GP process includes several selections such as execution and succession, mutation, and crossover processes. After the end of each round, the agent with lowest rank is killed. Next, three agents in the lowest rank are mutated. Finally, one child is generated with the crossover of agents with 1st and 2nd ranks and added to the group.

There are three mutation processes. For the first process (10%), one of the nodes on an agent's strategy tree is selected and its state is inverted. For the second process (80%), one of the edges is selected and its goal is attached to another node. For the third process (10%), a new node is added to the strategy tree. An orphan node created by the crossover process may be connected according to the mutation process. Figure 2 shows these processes.

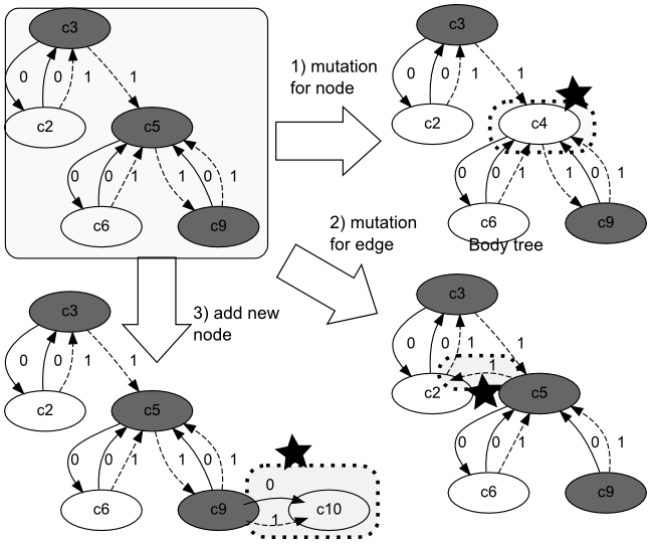

Figure 2. Mutation process on GP. Top-left automaton changed at black stars

In the crossover process, one of the pair (1st and 2nd ranks) is randomly selected as the body of the new agent, and its strategy tree is partially replaced with the randomly selected node of the other's strategy tree. A replacement tree is selected from the other of the pair's randomly selected nodes. The entire process is shown in Fig. 3. A black star on the body of the tree is replaced with a white star on the crossover tree. The cell number is replaced with a new and unused number when the number is already used on original tree. If orphan nodes and edges are generated according to the mutation and crossover processes, they are preserved for future mutations and crossovers.

Variation in the initialization phase is important for achieving good results from GP. The author used several simple strategies noted in previous studies as the seeds of this simulation. In the initialization phase, the system creates 50 random agents. Each agent has 1 or 2 states and each starting point and four edges are randomly assigned. As a result, the game will have 32 possible variations of agents. Some of the agents have the well-known strategies illustrated in Fig. 1.

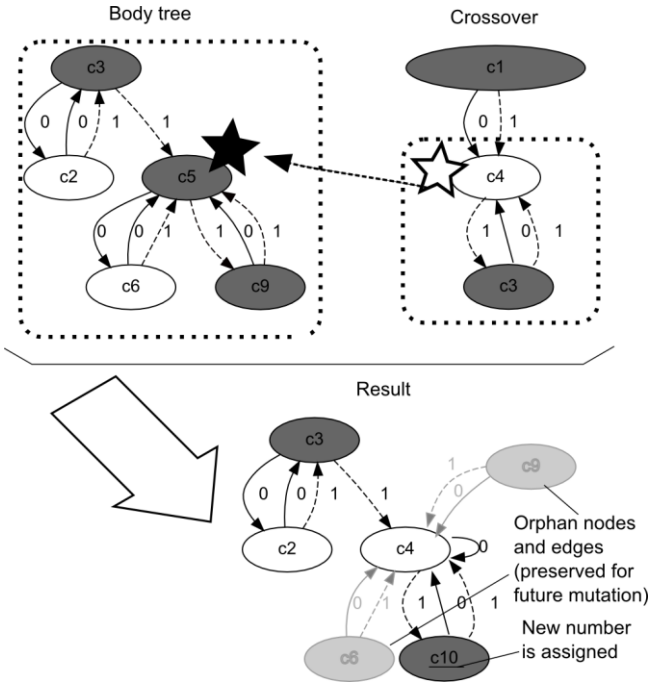

Figure 3. Crossover process.

2.5 Hypotheses

From the prediction discussed in Section 2.1 and setup discussed in the above sections, the author forms the following hypotheses.

1) The size of an agent's strategy tree increases in AMPD, which Byrne describes as an intelligence arms race [2]. Previous IPD simulations with GP show that IPD will converge in a mixture of several simple strategies. On the other hand, human-based simulations suggest that AMPD will increase the amount of automaton state [12] (There are no related studies on MMPD game, so it is required to find the result on simulation at same time).

2) The complexity of an agent's strategy tree also increases. The author calculated the average complexity of each agent in each

round by cyclomatic complexity [18]. In this simulation, cyclomatic complexity is simply calculated by subtracting connected nodes from unique connecting edges, because these automatons do not have an exit node. For example, the cyclomatic complexity of the believer and attacker in Fig. 1 is calculated as 0 because there are a node and one unique edge (c2->c2 or c1->c1). The complexity of GRIM is calculated as 1 because there are two nodes and three unique edges (c2->c2, c2->c1, and c1->c1). The complexities of TFT, aTFT, and Pavlov are calculated as 2 because there are two nodes and four unique edges (c2->c2, c2->c1, c1->c1, and c1->c2). The author wants to emphasize that the number of cyclomatic complexity does not directly refer to real intelligence. However, intelligent strategy requires several branches in it and it increases cyclomatic complexity. The author estimates that if an automaton has more cyclomatic complexity, it acquires the ability to achieve more complex behavior.

3. RESULTS

The maximum average score in IPD is 3 and maximum average scores in MMPD and AMPD is 2. In IPD (100 games), no agent acquired more than 5 nodes (10 edges) in any game. The final agent strategies were a mixture of TFT, believer, and GRIM, similar to previous simulations on IPD. Figure 4 shows the average amount of edges, average amount of used edges, and average score of IPD. The average score quickly increased to approximately 3 (achieved by C-C) during the first 50 rounds. In MMPD (100 games), agents in 92 games had less than 5 states, and agents in 8 games acquired more than 5 nodes and less than 10 nodes of automatons (10-45 edges). However, the author found that these cases were caused by only duplicated automaton states which is not used and did not really represent the complexity of agent strategies. Figure 5 shows the average amount of edges, average amount of used edges, and average score of MMPD. The average of all edges slightly increased in the 8 games. However, used edges did not increase in any case. The average score quickly increased to approximately 2 (achieved by continuous C-Cs or continuous C-D and D-C pairs) during the first 50 rounds.

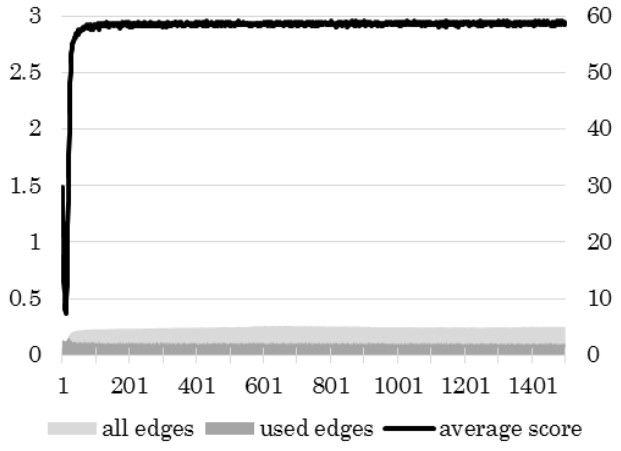

Figure 4. Average scores and edges on IPD. Left Y axis shows the score of the agents per each match. Right Y axis shows the average amount of edges in each automaton. Dark gray shows average used edges and light gray shows average all edges.

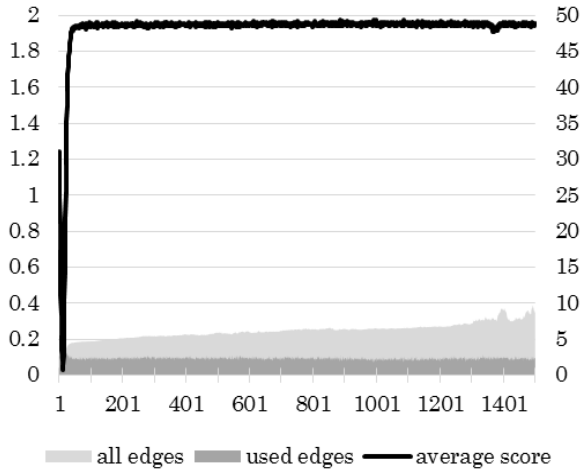

Figure 5. Average scores and edges on MMPD. Notations are same as in Fig. 4.

In AMPD (100 games), agents in 57 games had less than 5 nodes. In 56 cases in these games, almost all agents had the GRIM or a similar simple strategy and their final score was almost 1 (between 0.9 to 1.05) because if the field was occupied and locked by GRIM strategies, no agent could acquire more rewards from the field. In one of the 57 games, the field was occupied and locked by attackers and the final score was -1. On the other hand, agents in 43 games finally acquired averagely 378 edges (between 51 to 1525, standard deviation SD=438). Used edges also increased in AMPD. In the final state, averagely 27 edges were still used as a working strategy (between 18 to 42, SD=6). Figure 6 shows the average amount of edges, average amount of used edges, and average score of AMPD in the 43 games.

The author also evaluated how many nodes and edges were used in each match. An edge was counted as used when at least one of the 49 opponents forced to use this strategic route during each round. The results are also shown in Fig. 4-6 with different colors. More than 10 nodes (20 edges) were used in the final round in AMPD.

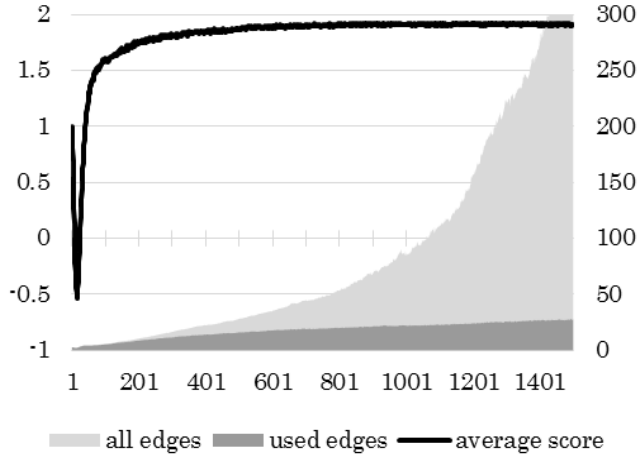

Figure 6. Average scores and edges on AMPD in 43 games. Notations are same as in Fig. 4.

The author also evaluated cyclomatic complexity in these 43 AMPD games. The complexity on the entire tree was averagely 172 (between 20 to 720, SD=202). The complexity on the used tree was averagely 7.0 (between 4.5 to 9.5, SD=1.2). Figure 7 shows how the average used edges and the average cyclomatic complexities on the used tree increased in each round. The black line shows the complexity on the used tree and the gray region shows the standard deviations. The increasing complexity of the tree suggests that each agent's strategy became more complex during rounds. The increasing speed of complexity on the used tree is relatively slow compared to the complexity on the entire tree. However, the complexity on the used tree became 4 and over in every agents in the final rounds. They became more complex than in the IPD and MMPD cases. In IPD and MMPD cases, the all cyclomatic complexities of used tree were 2 and fewer. This result means that no more complex strategies than simple strategies (as shown in Fig 1) were generated in IPD and MMPD cases.

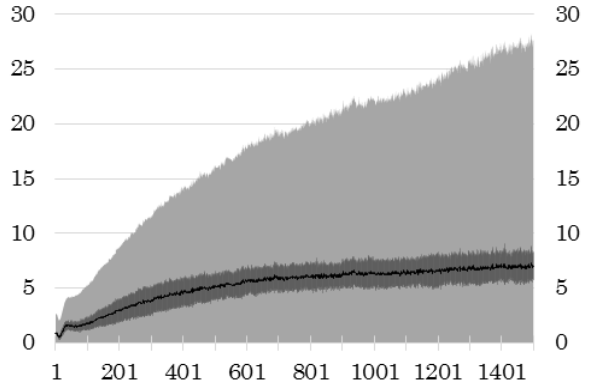

Figure 7. Average cyclomatic complexity on used tree. X axis shows rounds. Light gray region shows average used edges. Black line shows average complexity on used tree and dark gray region shows standard deviations.

We applied regression analysis between rounds and four valuables (average nodes of automatons, average used nodes of automatons, average cyclomatic complexity on the entire strategy tree, and average cyclomatic complexity on used strategy tree) in 43 games. All correlation coefficients were over 0.85 and p-value was under 0.001. These statistical results suggest that increases in the size of automatons and cyclomatic complexity are significant.

4. DISCUSSION

The author compares the three payoff scheme (IPD, MMPD, and AMPD) and give a detailed analysis using a sample of AMPD. The detailed analysis was conducted by referring to several samples from the game and evaluating the history of each agent. This evaluation method is based on biological analysis of the artificial life system Tierra [19]. The author selected several agents from the digital world as subjects and dissected them.

4.1 Comparison of IPD, MMPD, and AMPD

The quick convergence on IPD, lower amount of average edges in each agent's strategy, and the fact that each agent's strategy involves very simple automatons (believer, TFT, GRIM) suggest that our simulation results match previous studies on IPD [20][21]. The convergence speed and total amount of edges are also quick and similar to those in MMPD. In MMPD, continuous C-Cs (both players cooperate) and continuous C-D and D-C pairs (players cooperate and defect, and play the other hand next time) both result in the same reward. The results suggest that there is no strong evolutionary pressure to make automatons complex during the game. In IPD and MMPD, simpler strategies are more appropriate to survive. The results suggested that the intelligence arms race suggested in social brain hypothesis is not found in IPD and MMPD.

In half the trials of AMPD, the edges and nodes of the automaton gradually increased in each round, which is different from IPD and MMPD. Each agent's strategy finally acquired averagely 378 edges. The result that 27 edges were averagely used shows that these strategies are not just an unnecessary byproducts from the GP process. It is also important that the other unused edges might not be garbage as a result of GP because these surplus nodes may show the robustness for possible trials. This robustness is not required in IPD and MMPD. An increase in the size of automatons was observed only in AMPD. The results also suggest that MMPD is categorized the same as IPD for evolving intelligence. A gradual increase in the average number edges of automatons shown in Fig. 6 supports Hypothesis 1. All agents in 43 AMPD games quickly reached the CD/DC loop (longer loops such as CD/CD/DC/DC or CD/CD/CD/DC/DC/DC are theoretically possible as solutions, but not found in these simulations. This is because a long loop has more risk of exploitation from opponents.) It is important that even though they quickly reach an equilibrium in scores, as shown in Fig. 6 (which suggests that scores are almost 2 points during the first 300 rounds), the arms race of strategies continues as the game proceeds. This may suggest "the red queen effect" in biology in which the arms race continues after equilibrium [22], occurred in our simulations. Also, gradual increases in the average number of cyclomatic complexity on the entire tree and used tree suggest that increased nodes contribute to making complex strategies. This result supports Hypothesis 2.

These results are different from a previous study. Osawa et al. suggested that human-based simulation increases intelligence and is supported both from AMPD and refusal selection on trade by each agent. Different from their study, this computational simulation suggests that refusal selection is not mandatory for evolving intelligence. The author still agrees with their results that the AMPD payoff table is mandatory for achieving the evolution of automatons. Each strategy should be unique in AMPD for avoiding collision of the same strategies. The situation is close to Takano et al.'s "walking a road with avoiding collision" problem [23].

4.2 Analysis of AMPD in detail: Imitations and complexity in strategies

For conducting a micro-based analysis, the author focused on an example trial. We chose the 3rd trial as a well-rounded example, because the number of its final states is the median of that of the 43 cases. Figures 8, 9, and 10 show unique automatons that were top-ranked during the first 150 rounds. This transitional history of top-ranked automatons helps to reveal what kind of increase in strategies occurred.

Each number on the top of each figure shows the first round number when each automaton appeared. The author named these

top-rank automatons according to their behavior as shown on the right position of each round number. The numbers on the edge sides show the kind of hand (C=0, D=1) and how many agents (up to 49 opponents) selected this edge as a route. The edge becomes thicker if the opponent selected this route. Gray edges mean that they were not used in its round. Although gray lines are not used, they would suggest the ability of an agent to handle possible opponents.

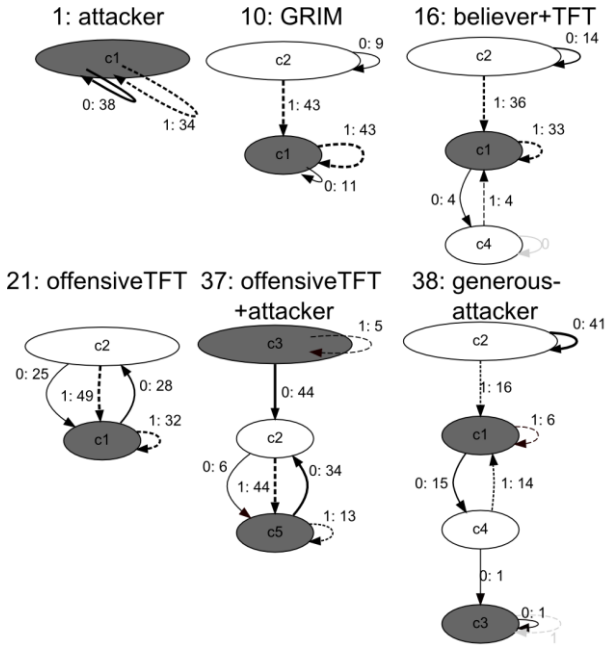

Figure 8. Top-ranked automatons from 3rd trial (1 to 38). Each number on the top of each figure shows round number. Numbers on side of edge denote type of hand (C=0 as solid lines, D=1 as dashed lines) and how many agents (up to 49 opponents, higher numbers expressed as thicker arrow) select this edge as a route. Gray edges mean that they are not used in each round.

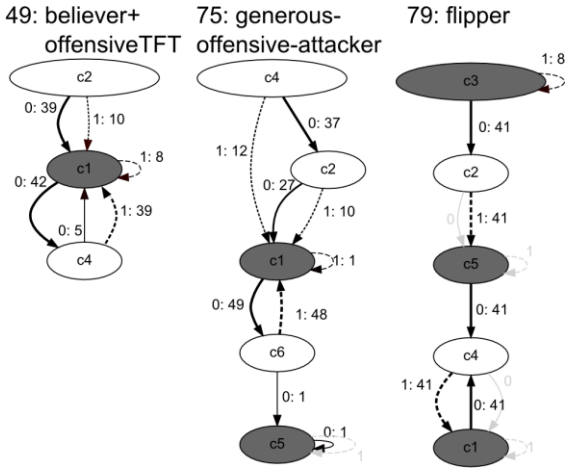

Figure 9. Top-ranked automatons from 3rd trial (49 to 79). Each number on the top of each figure shows round number and other notations are same as in Fig. 7.

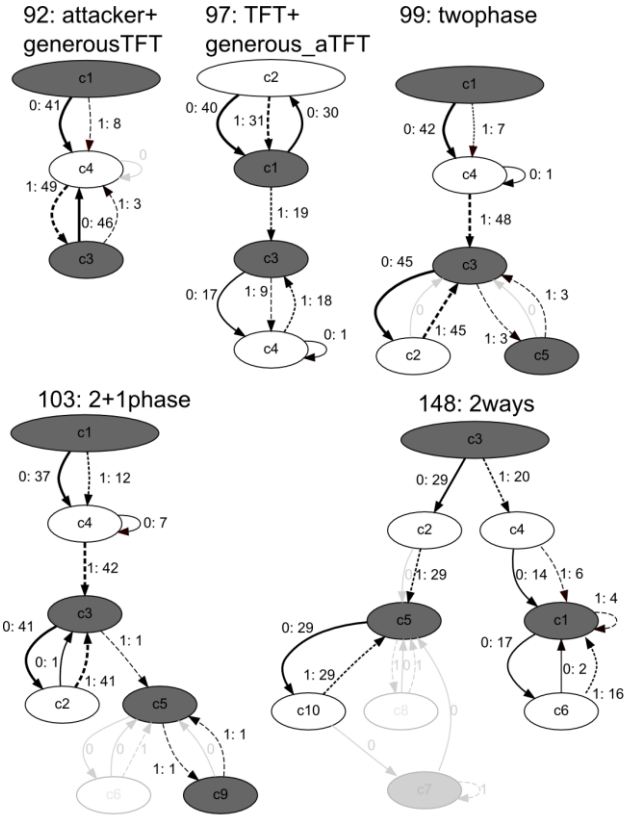

Figure 10. Top-ranked automatons from 3rd trial (92 to 148). Each number on the top of each figure shows round number and other notations are same as in Fig. 7.

With Figs. 8 to 10, we can confirm a general trend in which the edges and nodes of top-ranked automatons increase. Each agent acquired branches during generations and this increased the complexity of their strategies. For example, strategies before number 97 is described as a tree with no separated strategies (shown in Figs. 8, 9, and 10). These automatons are described as straight program and the program counter proceeds and backs on the line according to the opponent's reactions. On the other hand, strategies after 99 (shown in Fig. 10) include separated stages. The separated stages become more complex in later automatons. These automatons also reveal that all strategies in Figs. 8 and 9, except numbers 92 and 97, have a routine for guarding against attackers (whose hands are all D) in every branch of their strategies. This suggests these automatons acquire basic ability for avoiding exploiters.

Several automatons include basic strategies inside itself. Although the results were obtained from the crossover process in the simulations, the author can also interpret that these agents learned "opponent's model" inside their automatons for mimicking workable logic from other agents to improve their strategies. For example, offensiveTFT+attacker (number 37 in Fig. 8) mimics offensiveTFT (number 21 in Fig.8) inside itself (between c2 to c5). The "2ways" strategy (number 148) is impressive because the right tree (between c4 to c6) is the same as "believer+offensiveTFT" (number 49). This agent first attacks an opponent and selects the behavior according to the reactions. If

the opponent reacts as a defect, this agent selects the conventional "believer+offensiveTFT" method. If the opponent cooperates, the agent steps into a more offensive routine, as shown in Fig. 10 number 148's left branch. This branch is similar to the right branch. However, this branch can produce a continuous DCD for believers, yet it is not used in this round. This circular routine is also able to exploit more generous strategies than TFT such as TFFT, which endures one defect and defects the opponent if it accepts two defects. It is also curious that a circular routine – loop with multiple nodes – is generated after 148 generations. These circular routines are suitable for producing three or more states periodical "signals." It should be noted that the evolution process forces agents to acquire this ability.

Figure 11 includes the ratio between when the strategy started from cooperate (which is categorized as "good" by Axelrod [5]) and from defect. The ratio becomes stable between 0.4 to 0.6 in the final rounds. This result suggests that variations in automatons become half-and-half. During the rounds, each agent created a more appropriate environment for making continuous C-D and D-C pairs or continuous D-C and C-D pairs.

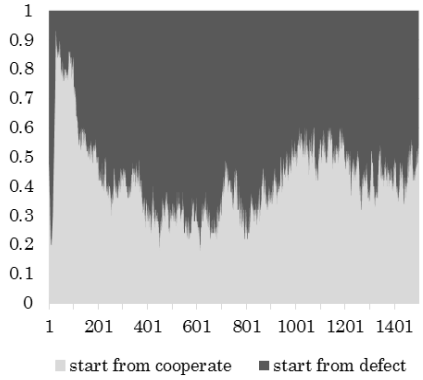

Figure 11. The ratio of cooperate and defect of start point. X axis shows rounds. Light gray region shows the ratio of agent who started from cooperate. Dark region shows the ratio of agent who started from defect.

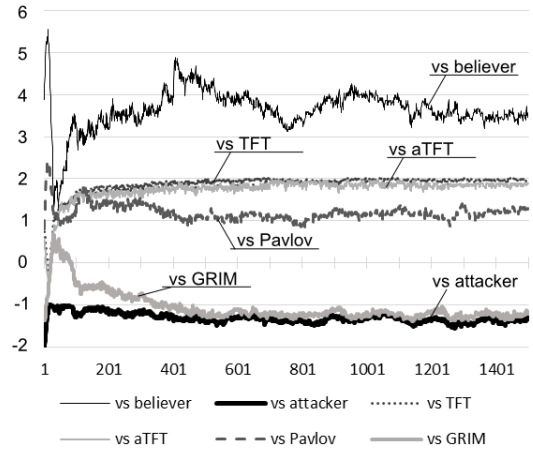

Figure 12. Average scores for simple strategies shown in Fig. 1. X axis shows rounds. Y axis shows the average score per each match.

The author evaluated how each selected agent behaved with the simple strategies (believer, attacker, TFT, aTFT, Pavlov, and GRIM) shown in Fig. 1. The results are shown in Fig. 12. Each agent acquired the appropriate behavior (CD/DC loop) for TFT and aTFT. They also acquired the ability to exploit believer and to suppress loss from the battle with an opponent and GRIM. Average achieved score from Pavlov was stable between 100 and 200. This is because Pavlov wants to sustain D if the player is C, (continuous C-Ds) and the player needs to play D and to make D-D pair in the game for breaking continuous C-Ds. Thus, maximum score is lower than the cases of TFT and aTFT. The battles for simple strategies suggests that the evolution process improves an agent's ability to handle the basic strategies of their opponents.

5. LIMITATIONS AND FUTURE WORKS

The author hypothesized that the intelligence of an agent and the amount of nodes and edges are proportional. From the simulation analysis, the author formed the three hypotheses given in Section 2.5. Although these hypotheses reasonably explain what occurred in the early rounds in our simulation, the increase in the size of automatons in the later rounds and increase in complexity does not directly mean increase in intelligence. Also, the author does not know how these simulations will continue for further rounds. For investigating intelligence and toughness of strategies in the simulations, further research is required to evaluate the complexity of an agent with more sophisticated methods (for example, evaluating evolved strategies using more complex strategies). Our GP does not have any limitation for length of automatons. The author wants to evaluate how the limitation of resource influences the evolution of intelligence in future.

The simulation and analysis results shown in Figs. 7 and 12, respectively, implicitly suggests the relationship between cyclomatic complexity and the ability for handling simple strategies. However, cyclomatic complexity does not approximately show this ability. Several different methods for evaluating the complexity of agent strategies are necessary. The author also would like to use qualitative analyses for imitating strategies and correlation between the similarity of agents and their scores.

The author wants to emphasize that there have been many studies for IPD regarding the evaluation of diversity and locality [6][8]. Our simulation gave a different viewpoint than previous life simulations, i.e., from the role of intelligence. Our simulations included only delayed rewards. However, there are several more complex dilemmas in actual trading. The author wants to evaluate how other factors will affect the intelligence of automatons for future work. Our simulations were affected by the AMPD setup from previous studies [7][12]. One of the significant differences of our evaluation is that our evaluation can not only be applied to an entire analysis, but also detail discussion on the selection of each automaton. The author founds some intelligent behavior (imitations, branches, categorization, and 3-states loops) are acquired during evolution.

Our findings revealed three important factors related to game theory, cognitive science, and artificial intelligence. From the game theory viewpoint, the possibility of mutual trading can be analyzed in the cheap talk game, which divides a trading game into an initialization phase and a main trading phase [24]. Our results suggest that identification of others and mutual cooperation can be achieved even without "cheap talk" by using

the reward itself. This finding may lead to multi-agent simulations becoming simpler as far as their requirements go. From the cognitive science viewpoint, the author finds that an intelligence arms race appears even with simpler rules. From the artificial intelligence viewpoint, the author found a general game rule to increase the complexity of autonomous agents. The results will contribute to creating the social behavior of non-player characters in video games [25] and social robots [26].

6. CONCLUSION

The author discussed the evolution of the intelligence arms race, represented as finite state automatons, under three distinct payoff schemes (IPD, MMPD, and AMPD). The author hypothesized that 1) an agent's strategy gradually increases, which Byrne described as an intelligence arms race [2], and 2) agent strategies in AMPD become complex. The results suggest that agents increase the complexity of their strategies and intelligence arms race occurs.

7. ACKNOWLEDGMENTS

This work was supported by the JST PRESTO program.

8. REFERENCES

[1] D. Premack and G. Woodruff, "Does the chimpanzee have a theory of mind?," *Behav. Brain Sci.*, vol. 1, no. 04, pp. 515–526, 1978.

[2] R. W. Byrne and A. Whiten, *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press, USA, 1989.

[3] S. Nolfi and D. Floreano, "Coevolving Predator and Prey Robots: Do 'Arms Races' Arise in Artificial Evolution?," *Artif. Life*, vol. 4, no. 4, pp. 311–335, Oct. 1998.

[4] M. V. Flinn, D. C. Geary, and C. V. Ward, "Ecological dominance, social competition, and coalitionary arms races," *Evol. Hum. Behav.*, vol. 26, no. 1, pp. 10–46, Jan. 2005.

[5] R. Axelrod, *The Evolution of Cooperation*. Basic Books, 1984.

[6] F. C. Santos, F. L. Pinheiro, T. Lenaerts, and J. M. Pacheco, "The role of diversity in the evolution of cooperation.," *J. Theor. Biol.*, vol. 299, pp. 88–96, Apr. 2012.

[7] P. J. Angeline, "An Alternate Interpretation of the Iterated Prisoner's Dilemma and the Evolution of Non-Mutual Cooperation," in *Proceedings of 4th artificial life conference*, 1994, pp. 353–358.

[8] R. Suzuki and T. Arita, "Evolutionary Analysis on Spatial Locality in the N-person Iterated Prisoner's Dilemma," *Proc. Inaug. Work. Artif. life*, pp. 105 – 114, 2001.

[9] E. van der Vaart and R. Verbrugge, "Agent-based models for animal cognition: a proposal and prototype," in *Autonomous Agents and Multi-Agent Systems*, 2008, pp. 1145–1152.

[10] S. C. Marsell, D. V Pynadath, and S. J. Read, "PsychSim : Agent-based modeling of social interactions and influence," in *Proceedings of the International Conference on Cognitive Modeling*, 2004, pp. 243–248.

[11] M. Si, S. C. Marsella, and D. V. Pynadath, "Modeling appraisal in theory of mind reasoning," *Auton. Agent. Multi. Agent. Syst.*, vol. 20, no. 1, pp. 14–31, May 2009.

[12] H. Osawa and M. Imai, "Evolution of Mutual Trust Protocol in Human-based Multi-Agent Simulation," in *12th European Conference on Artificial Life*, 2013, pp. 692–697.

[13] S. Le and R. Boyd, "Evolutionary dynamics of the continuous iterated prisoner's dilemma," *J. Theor. Biol.*, vol. 245, no. 2, pp. 258–67, Mar. 2007.

[14] M. Nowak and K. Sigmund, "Chaos and the evolution of cooperation.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 90, no. 11, pp. 5091–4, Jun. 1993.

[15] C. Wedekind and M. Milinski, "Human cooperation in the simultaneous and the alternating Prisoner's Dilemma: Pavlov versus Generous Tit-for-Tat.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 7, pp. 2686–9, Apr. 1996.

[16] D. Kraines and V. Kraines, "Evolution of Learning among Pavlov Strategies in a Competitive Environment with Noise," *J. Conflict Resolut.*, vol. 39, no. 3, pp. 439–466, Sep. 1995.

[17] R. Fisher and D. Shapiro, *Beyond Reason: Using Emotions as You Negotiate*. Viking Adult, 2005, p. 256.

[18] T. J. McCabe, "A Complexity Measure," *IEEE Trans. Softw. Eng.*, vol. SE-2, no. 4, pp. 308–320, 1976.

[19] T. S. Ray, "Evolution and optimization of digital organisms," in *Scientific Excellence in Supercomputing: The IBM 1990 Contest Prize Papers*, K. R. et al. Billingsley, Ed. The Baldwin Press, 1991, pp. 489–531.

[20] R. Axelrod and D. Dion, "The further evolution of cooperation.," *Science*, vol. 242, no. 4884, pp. 1385–90, Dec. 1988.

[21] H. Iba, "Evolutionary learning of communicating agents," *Inf. Sci. (Ny).*, vol. 108, no. 1, pp. 181–205, 1998.

[22] L. Van Valen, "A new evolutionary law," *Evol. Theory*, vol. 1, pp. 1–30, 1973.

[23] M. Takano and T. Arita, "Asymmetry between Even and Odd Levels of Recursion in a Theory of Mind," in *ALife X*, 2006, pp. 405–411.

[24] K. Wärneryd, "Evolutionary stability in unanimity games with cheap talk," *Econ. Lett.*, vol. 36, no. 4, pp. 375–378, Aug. 1991.

[25] M. Ochs, N. Sabouret, V. Corruble, U. Pierre, and C. Paris, "Simulation of the Dynamics of Non-Player Characters' Emotions and Social Relations in Games," *IEEE Trans. Comput. Intell. AI Games*, vol. 1, no. 4, pp. 281–297, 2009.

[26] M. Mataric, "Learning to Behave Socially," in *International Conference on Simulation of Adaptive Behavior*, 1994, pp. 453–462.