T. Sudo, K. Goto, Y. Nojima, and H. Ishibuchi, Effects of ensemble action selection with different usage of player's memory resource on the evolution of cooperative strategies for iterated prisoner's dilemma game," Proc. of 2015 IEEE Congress on Evolutionary Computation, pp. 1505-1512, Sendai, Japan, May 25-28, 2015.

Effects of Ensemble Action Selection with Different Usage of Player's Memory Resource on the Evolution of Cooperative Strategies for Iterated Prisoner's Dilemma Game

Takahiko Sudo, Kazushi Goto, Yusuke Nojima, and Hisao Ishibuchi
Department of Computer Science and Intelligent Systems
Graduate School of Engineering, Osaka Prefecture University
Sakai, Osaka 599-8531, Japan
{takahiko.sudo@ci., kazushi.goto@ci., nojima@, hisaoi@}cs.osakafu-u.ac.jp

***Abstract*—In our previous study, we proposed an ensemble action selection model where each player has multiple strategies with different memory length for the iterated prisoner's dilemma (IPD) game. An action was suggested by each strategy based on its memory about opponent's single, two or three actions. Majority vote was used for action selection. Under these settings, the evolution of cooperation was examined for various ensembles (i.e., various combinations of strategies). In this paper, we extend our ensemble model to a more general case where strategies have different memory usage. Each strategy of a player has a memory of opponent's and/or player's previous actions. For example, a memory of a strategy can be opponent's single and player's two actions. Another strategy's memory can be player's three actions. Various combinations of strategies for ensemble action selection are examined in this paper. It is shown through computational experiments that the use of ensemble action selection enhances the evolution of cooperation. It is also shown that no cooperation is evolved among strategies with no memory about opponent's actions. An interesting observation is that cooperation is evolved by players with the combination of the following three strategies: two strategies with no memory about opponent's actions, and a single strategy with a memory of both player's and opponent's actions.**

***Keywords*—iterated prisoner's dilemma game; ensemble action selection; memory resource allocation**

I. Introduction

The prisoner's dilemma (PD) game is a well-known two-player non-zero sum game. Its iterated version (IPD game: iterated prisoner's dilemma game) has been frequently used to study the evolution of cooperative strategies [1]-[3], [13]. In many studies on IPD games, each player has only a single strategy. However, our daily decision making is not always based on a single strategy. We often use different viewpoints (i.e., different strategies) to make a decision. Motivated by these discussions, we proposed in our previous study [4] an ensemble IPD model where each player had an ensemble of strategies for action selection. Majority vote was used to make a decision using multiple strategies. Each strategy had a memory of opponent's single, two or three actions (and no memory about player's own actions). That is, three different types of strategies were used in [4], each of which had memory length one, two or three. It was demonstrated that the use of the ensemble IPD model enhanced the evolution of cooperation among players in comparison with the use of a single strategy.

In [4], it was also demonstrated that memory length has a large effect on the evolution of cooperation in a noisy situation with a small error probability of action selection. Another important factor with respect to player's memory is memory usage [5]. For example, when the memory length is three, there exist the following four types of memory usage: opponent's three actions as in our previous study [4], opponent's two actions and player's single action, opponent's single action and player's two actions, and player's three actions. In this paper, we extend our ensemble IPD model to a more general case where strategies have different memory usage.

This paper is organized as follows. We first explain the IPD game and our ensemble IPD model in Section II. Next we explain our genetic algorithm for the evolution of ensembles of IPD game strategies in Section III. Then we report experimental results in Section IV. Finally, we conclude this paper in Section V.

II. IPD Game with Our Ensemble Model

A. IPD Game

In the IPD game, a player and its opponent choose either "C: Cooperation" or "D: Defection", simultaneously. We use a standard payoff matrix in Table I. When the player defects and its opponent cooperates, the player receives payoff 5 which is the highest payoff in Table I. When the player cooperates and its opponent defects, the player receives payoff 0 which is the lowest payoff in Table I. When both of them cooperate, each of them receives payoff 3. This is the maximum average payoff of the player and the opponent. When both of them defect, each of them receives payoff 1. This is the minimum average payoff of the player and the opponent. As shown in Table I, the player can receive a higher payoff by defecting independent of its opponent's action. That is, the defection is a rational action for both the player and the opponent in Table I. However, the choice of a rational action by the player and the opponent (i.e., mutual defection) leads to the minimum average payoff 1.

B. Ensemble Action Selection in the IPD Game

Each player usually has a single strategy in many studies on the IPD game [6]-[9]. However, in our ensemble IPD model in this paper, each player has three strategies as in our previous work [4]. Majority vote is used to choose an action using the suggestions by the three strategies as illustrated in Table II. For example, when the three strategies suggest actions "C", "D" and "C", "C" is selected as a player's action.

TABLE I. PAYOFF MATRIX OF OUR IPD GAME

Player's Action	Opponent's Action	
	C: Cooperation	D: Defection
C: Cooperation	Player Payoff: 3 Opponent Payoff: 3	Player Payoff: 0 Opponent Payoff: 5
D: Defection	Player Payoff: 5 Opponent Payoff: 0	Player Payoff: 1 Opponent Payoff: 1

TABLE II. ILLUSTRATION OF ENSEMBLE ACTION SELECTION

Player's action	D	D	D	C	D	C	C	C
Suggestion by Strategy 1	D	D	D	D	C	C	C	C
Suggestion by Strategy 2	D	D	C	C	D	D	C	C
Suggestion by Strategy 3	D	C	D	C	D	C	D	C

C. Strategies with Different Memory Usage

In this paper, strategies of each player have different memory usage whereas they have the same memory length (which is three in this paper). There exist four types of memory usage under this setting: (a) opponent's three actions (O3), (b) opponent's two actions and player's single action (O2-P1), (c) opponent's single action and player's two actions (O1-P2), and (d) player's three actions (P3). Strategies with each memory usage type are represented by the following binary strings as explained in Tables III-VI: (a) 15-bit string, (b) 13-bit string, (c) 13-bit string, and (d) 15-bit string.

Table III explains how a 15-bit strategy of type (a) determines an action using its memory of opponent's three actions. This strategy is called the O3 memory strategy. The first bit x_1 determines the player's action in the first round using no information about the opponent's previous actions. The second bit x_2 and third bit x_3 determine the player's action in the second round based on the opponent's action in the first round (D => x_2, C => x_3). The next four bits "$x_4x_5x_6x_7$" determine the player's action in the third round based on the opponent's actions in the first and second rounds (DD => x_4, CD => x_5, DC => x_6, CC => x_7). The other eight bits "$x_8x_9x_{10}x_{11}x_{12}x_{13}x_{14}x_{15}$" determine the player's action in the tth round ($t > 2$). For example, opponent's actions were "D" in the (t–3)th round, "D" in the (t–2)th round and "D" in the (t–1)th round, x_8 is used to determine the player's action in the tth round.

Table IV explains how a 13-bit strategy of type (b) determines an action using its memory of opponent's two actions and player's single action, which is called the O2-P1 memory strategy. The first bit x_1 determines the player's action in the first round. The next four bits "$x_2x_3x_4x_5$" determine the player's action in the second round using the opponent's action and the player's action in the first round. The other eight bits "$x_6x_7x_8x_9x_{10}x_{11}x_{12}x_{13}$" determine the player's action in the tth round ($t > 2$) using the opponent's actions in the (t–2)th and (t–1)th rounds and the player's action in the (t–1)th round.

In the same manner, Table V shows how a 13-bit strategy of type (c) determines an action using its memory of opponent's single action and player's two actions, which is called the O1-P2 memory strategy. Table VI explains a 15-bit strategy of type (d) with a memory of player's own three actions, which is called the P3 memory strategy.

In our computational experiments, the PD game between the same pair of a player and an opponent is iterated 100 times (i.e., 100 rounds). That is, the number of rounds of the PD game in a single execution of the IPD game is 100.

We use an error probability P_{Error} in action selection in each round of the PD game. It is assumed that each player (i.e., each of a player and its opponent) chooses an action different from the determined one by an ensemble of its three strategies with a pre-specified error probability P_{Error}. We examine two settings of the error probability: $P_{\text{Error}} = 0$ and $P_{\text{Error}} = 0.01$. The error probability 0 means that each player always chooses the action determined by its three strategies (i.e., noise-free setting). The error probability 0.01 means that each player chooses an action different from the determined one with the probability 0.01 (i.e., noisy setting). That is, each player makes an error once in 100 rounds on average.

TABLE III. TYPE (a) STRATEGY WITH THE O3 MEMORY

Opponent's (t-3)	-	-	-	-	-	-	-	D	D	D	D	C	C	C	C
Opponent's (t-2)	-	-	-	D	D	C	C	D	D	C	C	D	D	C	C
Opponent's (t-1)	-	D	C	D	C	D	C	D	C	D	C	D	C	D	C
Player's action	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}

TABLE IV. TYPE (b) STRATEGY WITH THE O2-P1 MEMORY

Opponent's (t-2)	-	-	-	-	-	D	D	D	D	C	C	C	C
Opponent's (t-1)	-	D	D	C	C	D	D	C	C	D	D	C	C
Player's (t-1)	-	D	C	D	C	D	C	D	C	D	C	D	C
Player's action	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}

TABLE V. TYPE (c) STRATEGY WITH THE O1-P2 MEMORY

Player's (t-2)	-	-	-	-	-	D	D	D	D	C	C	C	C
Opponent's (t-1)	-	D	D	C	C	D	D	C	C	D	D	C	C
Player's (t-1)	-	D	C	D	C	D	C	D	C	D	C	D	C
Player's action	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}

TABLE VI. TYPE (d) STRATEGY WITH THE P3 MEMORY

Player's (t-3)	-	-	-	-	-	-	-	D	D	D	D	C	C	C	C
Player's (t-2)	-	-	-	D	D	C	C	D	D	C	C	D	D	C	C
Player's (t-1)	-	D	C	D	C	D	C	D	C	D	C	D	C	D	C
Player's action	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}

D. IPD Models

When we perform computational experiments under the standard non-ensemble framework, one of the four types of strategies (a)-(d) in Tables III-VI is chosen and assigned to all players. In our ensemble IPD model, a combination of three strategies is chosen as an ensemble from the four types (a)-(d) of strategies. The selected three types of strategies are assigned to all players. In this paper, we examine the 20 combinations of three types of strategies in Table VII. For example, ensemble A has three strategies of type (a) in Table III. Ensemble B has two strategies of type (a) and a single strategy of type (b) in Tables III and IV. One of those combinations is selected and assigned to all players.

In our computational experiments, each strategy assigned to each player is initialized randomly by specifying 0 (i.e., D) and 1 (i.e., C) with the same probability to each value in the binary string of the corresponding strategy.

TABLE VII. COMBINATIONS OF THREE BINARY STRINGS OF DIFFERENT MEMORY

Ensemble Model	A	B	C	D	E	F	G	H	I	J
Strategy 1	a	a	a	a	a	a	a	a	a	a
Strategy 2	a	a	a	a	b	b	b	c	c	d
Strategy 3	a	b	c	d	b	c	d	c	d	d
Ensemble Model	K	L	M	N	O	P	Q	R	S	T
Strategy 1	b	b	b	b	b	b	c	c	c	d
Strategy 2	b	b	b	c	c	d	c	c	d	d
Strategy 3	b	c	d	c	d	d	c	d	d	d

III. STRATEGY EVOLUTION

A. Spatial Structures

It is well-known that a spatial structure of players has a large effect on the evolution of cooperation in the IPD game. Various spatial structures (e.g., a grid-world and a network) have been examined [7]-[12]. In this paper, we examine two settings: a population model with no spatial structure and a two-dimensional grid-world. In the non-spatial setting, there is no neighborhood structure. That is, each player can be viewed as having all the other players as its neighbors. The IPD game can be played among any players. In the grid-world setting, each player has its neighbors specified by a neighborhood structure. We assume the torus structure in the grid-world. A 7×7 grid-world is shown in Fig. 1 where four neighbors are defined for each player by the von Neumann neighborhood.

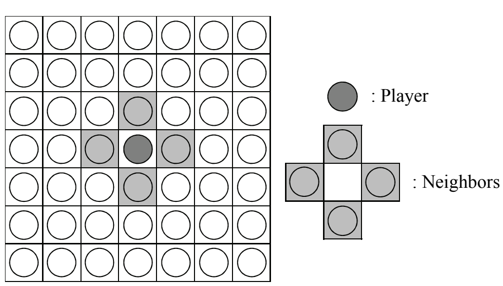


Fig. 1. Illustration of a grid-world with the torus structure.

In our computational experiments, each player plays the IPD game against a pre-specified number of opponents (four opponents in this paper). The opponents of each player are randomly selected from its neighbors without replacement. The IPD game is iterated between a player and its opponent for a pre-specified number of rounds (100 rounds in this paper). The fitness of each player is the average payoff obtained by the IPD game against its all opponents.

B. Evolution of Ensembles

To generate new strategies in the standard single-strategy model, each player chooses a pair of players as parents by binary tournament selection with replacement. In the non-spatial setting, the parents are selected from all players (including the player itself). In the grid-world setting, they are selected from the neighbors and the player itself. For example, two parents are selected from the four neighbors and the player in Fig. 1 using binary tournament selection with replacement based on the average payoff of the player and each neighbor.

In the standard single-strategy model, we apply one-point crossover to the strategies of the selected parents to generate two strategies. The crossover probability is specified as 1.0 in this paper. One of the generated two strategies is randomly selected as an offspring. Bit-flip mutation is applied to the selected offspring with the probability $1/NL$, where N and L are the population size (i.e., the number of players) and the string length, respectively. The mutation is applied to a single bit of a single strategy among N strategies of length L on average.

In our ensemble model where each player has three strategies, we select two parents for each player in the same manner as in the standard single-strategy model. Then, one-point crossover and bit-flip mutation are applied to the first strategies of the selected two parents in the same manner as in the standard single-strategy model to generate a new strategy as the first strategy of the player. In this manner, the second and third strategies are generated by the second and third strategies of the selected two parents, respectively.

After new strategies are generated for all players, their current strategies are replaced with the new strategies in a synchronized manner. The number of generations is specified as 1000 and used as the termination condition.

IV. COMPUTATIONAL EXPERIMENTS

A. Experimental Setting

We examine 24 different settings with respect to player's strategies, which include four settings of the single-strategy model and 20 settings of the ensemble model. Each setting of the single-strategy model has one of the four types of strategies with different memory usage (i.e., types (a)-(d) in Tables III-VI in Section II: O3, O2-P1, O1-P2 and P3 memories). The 20 settings of the ensemble model are shown in Table VII in Section II. Each ensemble setting has a different combination of strategy types in an ensemble. In our computational experiments, the same setting of strategies is assigned to all players in each setting.

We use a 32 × 32 grid-world with four neighbors in the von Neumann neighborhood. The number of players is the same as the number of cells: 32 × 32 = 1024. We also use a non-spatial population model. Two settings of the error probability are examined: $P_{\mathrm{Error}} = 0$ in the noise-free case and $P_{\mathrm{Error}} = 0.01$ in the noisy case. Thus 2 × 2 combinations of the spatial structure and the error probability are examined for each of the 24 settings of player's strategies. As a result, the evolution of cooperation is examined for 2 × 2 × 24 = 96 different settings in this paper.

The number of generations for strategy evolution is specified as 1000. The average payoff is calculated over 1000 runs with 1000 generations for each of the 96 settings. When the average payoff is close to 3.0, we can say that the mutual cooperation is evolved for almost all players. If the average payoff is close to 1.0, we can see that the mutual defection is evolved for almost all players.

B. Results by the Standard Single-Strategy Models

Experimental results by the four settings of the standard single-strategy model are shown in Figs. 2-5. Fig. 2 is the results of the standard single-strategy model with a memory of opponent's three actions in the previous three rounds (i.e., the O3 memory strategy). In the noise-free setting (i.e., $P_{\mathrm{Error}} = 0$ in Fig. 2), the average payoff rapidly increased close to 3.0 within 100 generations in both the spatial grid-world model in Fig. 2 (a) and the non-spatial population model in Fig. 2 (b). Fig. 2 also shows that the use of the error probability 0.01 (i.e., the noisy setting with $P_{\mathrm{Error}} = 0.01$) decreased the average payoff. The negative effect of the error probability was more severe in the non-spatial setting in Fig. 2 (b) than the spatial setting in Fig. 2 (a). This observation is consistent with results reported in the literature that the spatial structure enhanced the evolution of cooperation.

Fig. 3 is the results of the standard single-strategy model with a memory of opponent's two actions in the previous two rounds and the player's single action in the previous single round (i.e., the O2-P1 memory strategy). In the noisy setting, the average payoff was much lower than Fig. 2. This shows that the strategy with a long memory about the opponent's actions is beneficial in the noisy setting.

Fig. 4 is the results of the standard single-strategy model with a memory of opponent's action in the previous round and the player's actions in the previous two rounds (i.e., the O1-P2 memory strategy). In the case of the grid-world, high average payoff was obtained as in Fig. 2 in both the noise-free and noisy settings. An interesting observation in Fig. 4 (b) in the non-spatial population model is that higher average payoff was obtained from the noisy setting than the noise-free setting. Further investigation may be needed to explain why those results were obtained from the non-spatial population model with the O1-P2 memory. A partial explanation may be as follows. This model can distinguish the two cases of opponent's defection: one is after player's defection and the other is after player's cooperation. That is, this model can differentiate the reciprocal defection and the other defection. Thus it may have a potential ability to handle the reciprocal defection to player's defection by error in a special manner in order to decrease the negative effect of the noisy setting. All the other O3, O2-P1 and P3 memory strategies cannot distinguish the two cases of defection.

Fig. 5 is the results of the standard single-strategy model with a memory of the player's actions in the previous three rounds (the P3 memory strategy). No cooperation was evolved by the players with no memory about opponent's actions.

C. Results by the Ensemble Models

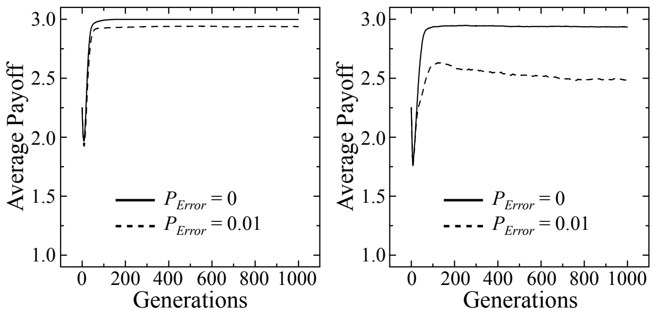
Figs. 6-9 show the results of three settings of the ensemble model where each setting has three strategies of the same memory type. That is, we can compare Figs. 6-9 with Figs. 2-5 to examine the effect of using three strategies of the same type on action selection. In the grid-world setting, the difference in the results between the standard single-strategy model and the ensemble model is not clear. However, in the non-spatial population setting, we can observe some differences between Figs. 6-8 and Figs. 2-4 (except for the case with the P3 memory strategies).

Lower average payoff was obtained in the noise-free setting in Fig. 6 (b) than in Fig. 2 (b). That is, we can observe a negative effect on the average payoff of using three O3 memory strategies instead of a single strategy of the same type. However, from the comparison between Fig. 3 (b) and Fig. 7 (b), we can observe a positive effect of using three O2-P1 memory strategies instead of a single strategy of the same type for the noisy setting. A positive effect of ensemble action selection is also observed from the comparison between Fig. 4 (b) and Fig. 8 (b) for the noisy setting.

In Figs. 10-12, we show the results of ensemble action selection by two O3 memory strategies and another strategy of a different type. That is, these figures can be viewed as the results after replacing one of the three O3 strategies in Fig. 6 with another strategy of a different type. In the case of the noise-free setting in the non-spatial population model, higher average payoff was obtained in Fig. 10 (b) than in Fig. 6 (b) by replacing one of three O3 memory strategies in Fig. 6 (b) with an O2-P1 memory strategy. We can observe a similar positive effect by replacing one of three O3 memory strategies with a different type from the comparison between Fig. 6 (b) and Fig. 11 (b). These observations suggest that the use of different types of strategies for ensemble action selection may have positive effects on the evolution of cooperation.

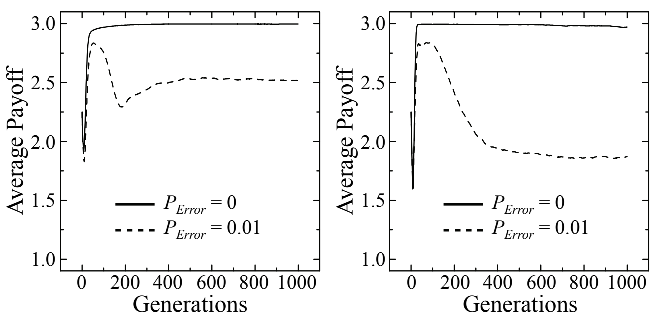
In Figs. 13-15, we show the results by ensembles of three strategies including two O2-P1 memory strategies. These figures are to examine the effect of replacing one of three O2-P1 memory strategies in Fig. 7 with a different type of strategy. From the comparison of Figs. 13-15 with Fig. 7, we can see that the replacement of one of three O2-P1 strategies with a different type of strategy increased the average payoff for the noisy setting in both the spatial and non-spatial settings.

In Figs. 16-18, we show the results by ensembles of three strategies including two O1-P2 memory strategies. These figures show the effect of replacing one of three O1-P2 memory strategies in Fig. 8 with a different type of strategy.



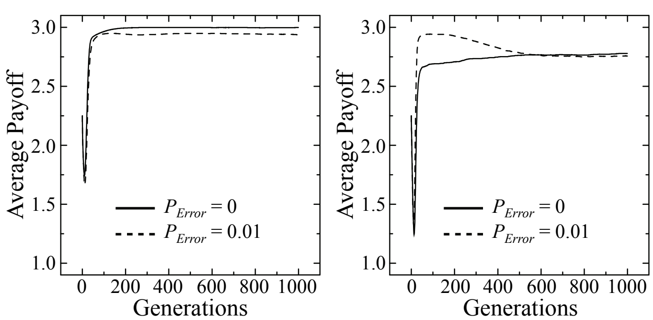
(a) Grid-world. (b) Non-spatial.

Fig. 2. Single-strategy model with the O3 memory.



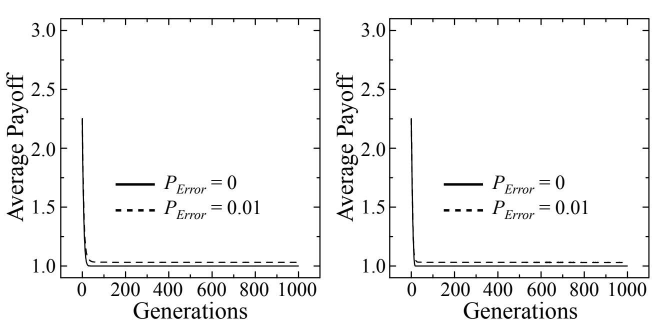
(a) Grid-world. (b) Non-spatial.

Fig. 3. Single-strategy model with the O2-P1 memory.



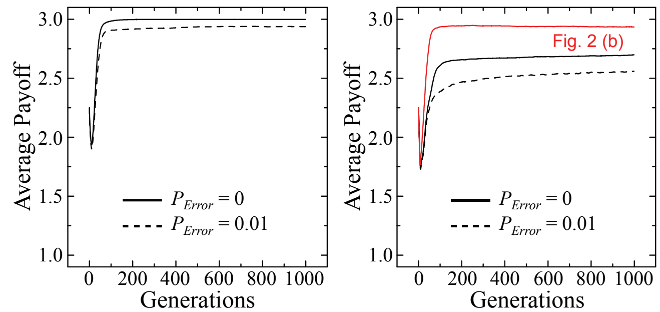
(a) Grid-world. (b) Non-spatial.

Fig. 4. Single-strategy model with the O1-P2 memory.



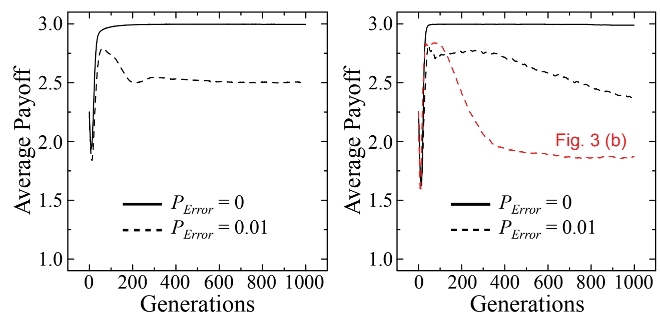
(a) Grid-world. (b) Non-spatial.

Fig. 5. Single-strategy model with the P3 memory.



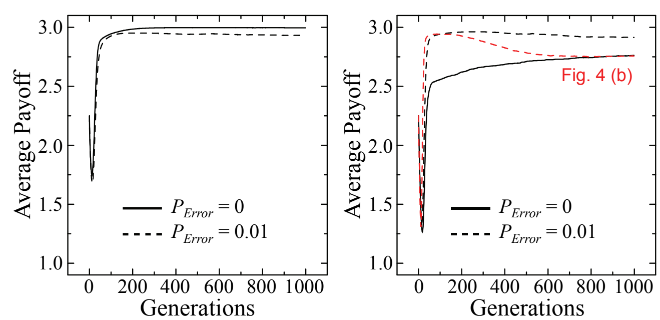
(a) Grid-world. (b) Non-spatial.

Fig. 6. Ensemble model of three strategies with the O3 memory.



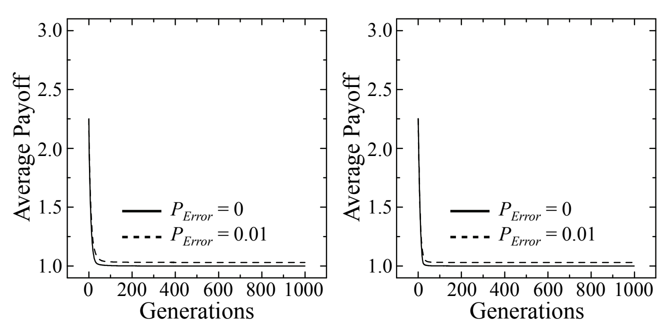
(a) Grid-world. (b) Non-spatial.

Fig. 7. Ensemble model of three strategies with the O2-P1 memory.



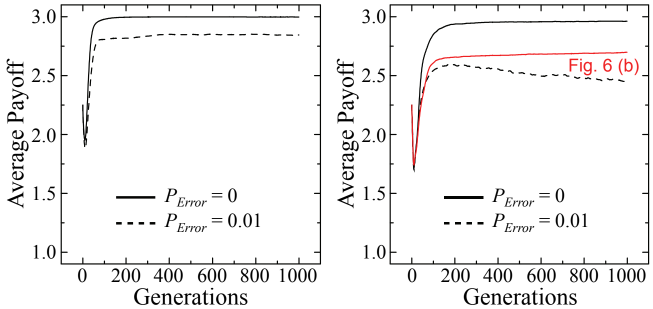
(a) Grid-world. (b) Non-spatial.

Fig. 8. Ensemble model of three strategies with the O1-P2 memory.



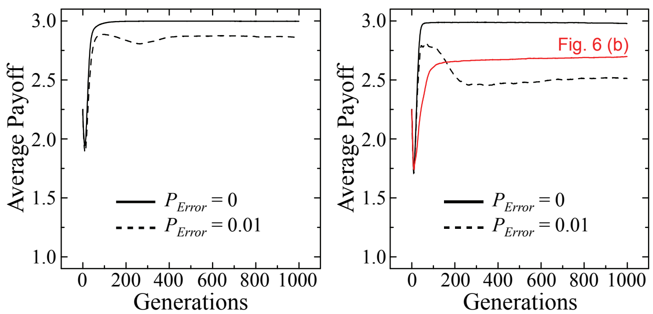
(a) Grid-world. (b) Non-spatial.

Fig. 9. Ensemble model of three strategies with the P3 memory.



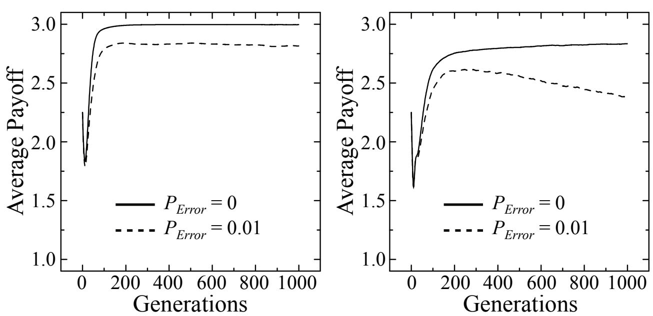
(a) Grid-world. (b) Non-spatial.

Fig. 10. Ensemble model with two O3 and one O2-P1 memory.



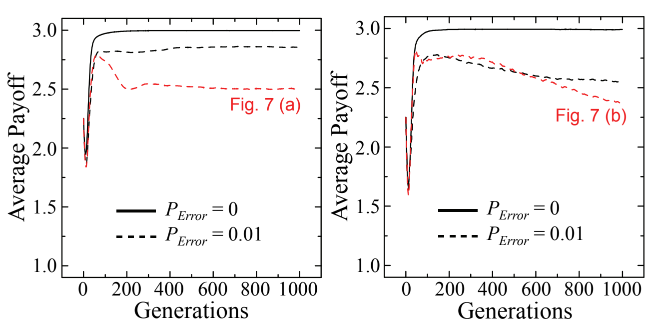
(a) Grid-world. (b) Non-spatial.

Fig. 11. Ensemble model with two O3 and one O1-P2 memory.



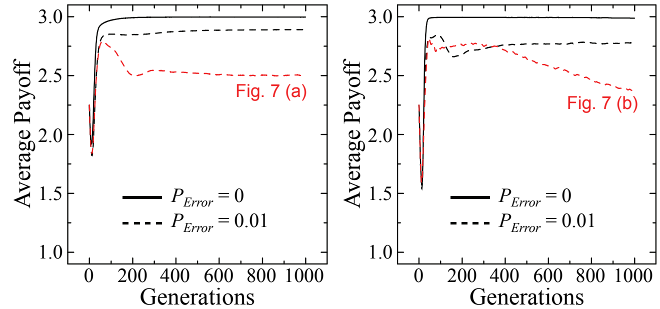
(a) Grid-world. (b) Non-spatial.

Fig. 12. Ensemble model with two O3 and one P3 memory.



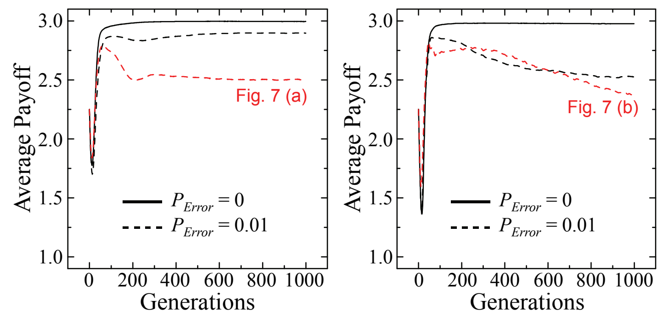
(a) Grid-world. (b) Non-spatial.

Fig. 13. Ensemble model with two O2-P1 and one O3 memory.



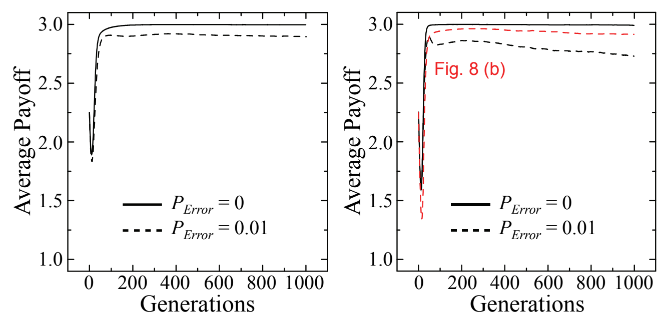
(a) Grid-world. (b) Non-spatial.

Fig. 14. Ensemble model with two O2-P1 and one O1-P2 memory.



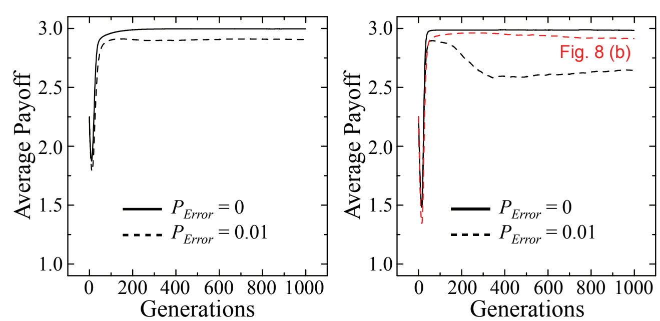
(a) Grid-world. (b) Non-spatial.

Fig. 15. Ensemble model with two O2-P1 and one P3 memory.



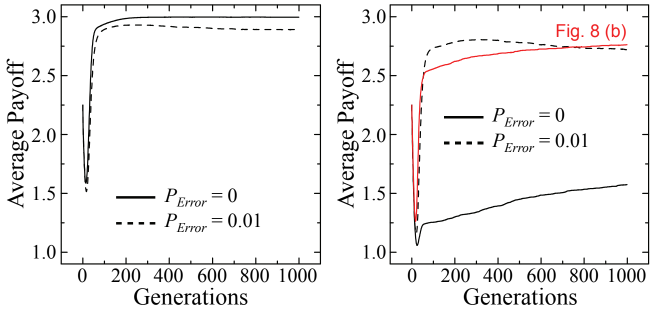
(a) Grid-world. (b) Non-spatial.

Fig. 16. Ensemble model with two O1-P2 and one O3 memory.



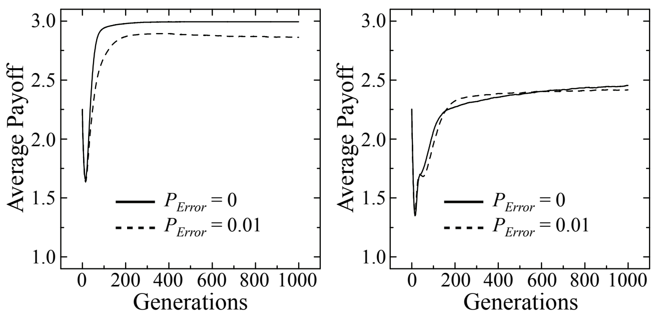
(a) Grid-world. (b) Non-spatial.

Fig. 17. Ensemble model with two O1-P2 and one O2-P1 memory.



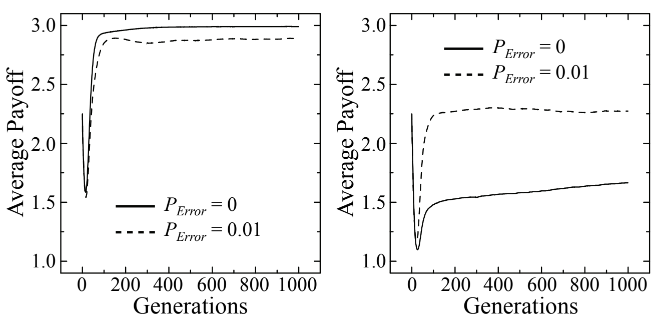
(a) Grid-world. (b) Non-spatial.

Fig. 18. Ensemble model with two O1-P2 and one P3 memory.



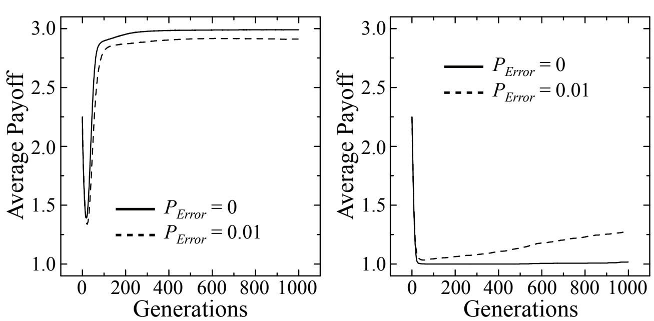
(a) Grid-world. (b) Non-spatial.

Fig. 19. Ensemble model with two P3 and one O3 memory.



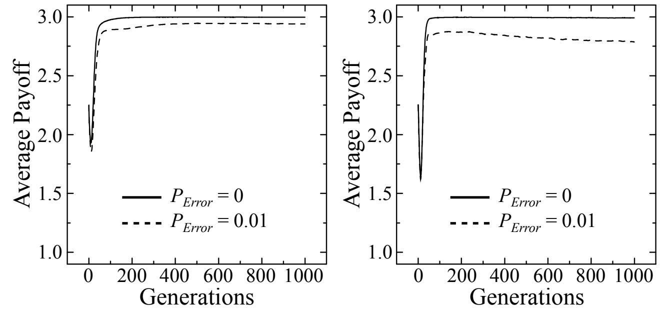
(a) Grid-world. (b) Non-spatial.

Fig. 20. Ensemble model with two P3 and one O2-P1 memory.



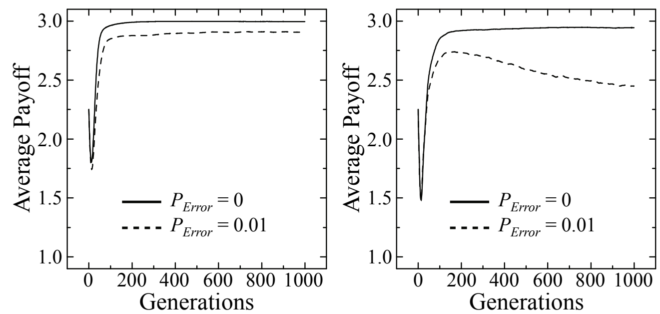
(a) Grid-world. (b) Non-spatial.

Fig. 21. Ensemble model with two P3 and one O1-P2 memory.



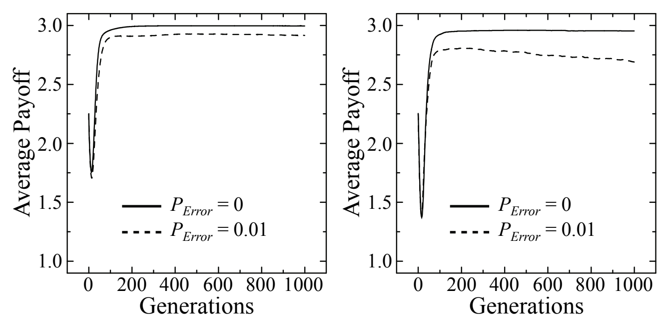
(a) Grid-world. (b) Non-spatial.

Fig. 22. Ensemble model with O3, O2-P1 and O1-P2 memory.



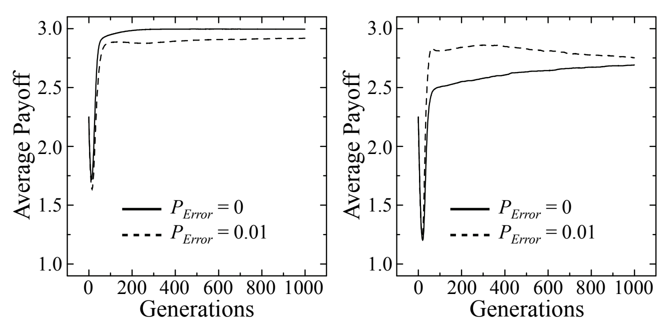
(a) Grid-world. (b) Non-spatial.

Fig. 23. Ensemble model with O3, O2-P1 and P3 memory.



(a) Grid-world. (b) Non-spatial.

Fig. 24. Ensemble model with O3, O1-P2 and P3 memory.



(a) Grid-world. (b) Non-spatial.

Fig. 25. Ensemble model with O2-P1, O1-P2 and P3 memory.

The strange behavior in Fig. 8 (b) disappeared in Fig. 16 (b) and Fig. 17 (b). That is, higher average payoff was obtained from the noise-free setting in Fig. 16 (b) and Fig. 17 (b) whereas higher average payoff was obtained from the noisy setting in Fig. 8 (b). Comparison between Fig. 8 (b) and Fig. 18 (b) shows a large effect of a single P3 memory strategy on the average payoff in the noise-free setting. It is also interesting to observe in Fig. 18 (b) that much higher average payoff was obtained from the noisy setting than the noise-free setting.

Figs. 19-21 show the results by ensembles of two P3 memory strategies and a different type of strategy. As shown in Fig. 5 and Fig. 9, no cooperation was evolved among players with one or three P3 memory strategies. Moreover, the inclusion of a single P3 memory strategy had a severely bad effect on the average payoff in Fig. 18 (b). From these observations, we expected no evolution of cooperation among players with two P3 memory strategies and a single different type strategy. However, we obtained high average payoff close to 3 when we used the grid-world for both the noise-free and noisy settings in Fig. 19 (a), Fig. 20 (a) and Fig. 21 (a). Much more investigation may be needed to explain this counter-intuitive observation from the ensembles with two P3 memory strategies and another strategy in Figs. 19-21.

In Figs. 22-25, we show the results by ensembles with three different types of strategies. The highest average payoff was obtained in Fig. 22 among Figs. 2-25. Moreover, the negative effect of the P3 memory model is not severe in Figs. 23-25 (if compared with Fig. 18). These observations from Figs. 22-25 may suggest that ensembles of different types of strategies enhance the evolution of cooperation.

V. CONCLUSION

In this paper, we examined the evolution of cooperation among players in the IPD game using the ensemble model where majority vote was used for action selection by each player with an ensemble of three strategies with different memory usage. Computational experiments were performed under two settings of errors in action selection (i.e., two settings of the error probability: 0 and 0.01) and two settings of spatial structures (i.e., a grid-world model and a non-spatial population model). Four types of memory usage of a length 3 memory were used in this paper (i.e., O3, O2-P1, O1-P2 and P3). With respect to the effect of the spatial structure, our observation was consistent to reported results in the literature: the use of spatial structures enhanced the evolution of cooperation. With respect to the effect of noise (i.e., errors in action selection), its positive effect was observed in some of our experiments whereas its negative effect was observed in most experiments. High average payoff was obtained by ensembles of three strategies of different types. An interesting observation was that the replacement of one of three P3 memory strategies with no memory about opponent's actions with a different type of strategy drastically changed the results from mutual defection to mutual cooperation. Further investments are needed to explain such an interesting observation. In this paper, the memory length of each player is defined as three. We also need to examine the case where the memory length is different from three as our future work.

REFERENCES

[1] R. Axelrod, "The evolution of strategies in the iterated prisoner's dilemma," in L. Davis (ed.), *Genetic Algorithms and Simulated Annealing*, Morgan Kaufmann, pp. 32-41, 1987.

[2] D. B. Fogel, "Evolving behaviors in the iterated prisoner's dilemma," *Evolutionary Computation*, vol. 1, no. 1, pp. 77-97, 1993.

[3] G. Kendall, X. Yao, and S. Y. Chong (eds.), *The Iterated Prisoners' Dilemma: 20 Years On*, World Scientific, Singapore, 2007.

[4] T. Sudo, Y. Nojima, and H. Ishibuchi, "Effects of ensemble action selection on the evolution of iterated prisoner's dilemma game strategies," *Proc. of 2014 IEEE Congress on Evolutionary Computation*, pp. 1195-1201, Beijing, China, 2014.

[5] D. Ashlock and E. Y. Kim, "The impact of varying resources available to iterated prisoner's dilemma agents," *Proc. of 2013 IEEE Symposium on Foundations of Computational Intelligence*, pp. 60-67, Wahington DC, USA, 2013.

[6] D. Ashlock, E. Y. Kim, and N. Leahy, "Understanding representational sensitivity in the iterated prisoner's dilemma with fingerprints," *IEEE Trans. on Systems, Man, and Cybernetics: Part C*, vol. 36, no. 4, pp. 464-475, July 2006.

[7] H. Ishibuchi, K. Hoshino, and Y. Nojima, "Evolution of strategies in a spatial IPD Game with a number of different representation schemes," *Proc. of 2012 IEEE Congress on Evolutionary Computation*, pp. 808-815, Brisbane, June 10-15, 2012.

[8] H. Ishibuchi, T. Sudo, K. Hoshino, and Y. Nojima, "Evolution of cooperative strategies for iterated prisoner's dilemma on networks," *Proc. of 5th International Conference on Computational Aspects of Social Networks*, pp. 32-37, Fargo, USA, August 12-14, 2013.

[9] D. A. Ashlock, "Cooperation in prisoner's dilemma on graphs," *Proc. of 2007 IEEE Symposium on Computational Intelligence and Games*, pp. 48-55, Honolulu, April 1-5, 2007.

[10] K. M. Bryden, D. A. Ashlock, S. Corns, and S. J. Willson, "Graph-based evolutionary algorithms," *IEEE Trans. on Evolutionary Computation*, vol. 10, no. 5, pp. 550-567, October 2006.

[11] X.-H. Deng, Y. Liu, and Z.-G. Chen, "Memory-based evolutionary game on small-world network with tunable heterogeneity," *Physica A*, vol. 389, no. 22, pp. 5173-5181, 2010.

[12] M. Mejia, N. Peña, J. L. Muñoz, O. Esparza, and M. Alzate, "A game theoretic trust model for on-line distributed evolution of cooperation in MANETs," *Journal of Network and Computer Applications*, vol. 34, no. 1, pp. 39-51, 2011.

[13] M. A. Nowak, "Five rules for the evolution of cooperation," *Science*, vol. 314, no. 5805, pp. 1560-1563, 2006.