

Interaction between Evolution and Learning in a Population of Globally or Locally Interacting Agents

Reiji SUZUKI and Takaya ARITA

Graduate School of Human Informatics, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, JAPAN
E-mail: {reiji, ari}@info.human.nagoya-u.ac.jp

Abstract

The Baldwin effect is known as one of the interactions between evolution and learning. In this paper we consider the Baldwin effect in dynamic environments, especially when there is no explicit optimal solution through generations and it depends only on the interaction among agents. We adopted the iterated Prisoner's Dilemma as a dynamic environment, introduced phenotypic plasticity into strategies, and conducted computational experiments on the models with/without spatial locality. In the non-spatial experiments, the Baldwin effect was observed as follows: First, strategies with enough plasticity spread, which caused a shift from defect-oriented population to cooperative population. Second, these strategies were replaced by a strategy with a modest amount of plasticity. Subsequently, we expanded this model in a two-dimensional space where each agent plays games only with its neighbors. Based on the results of these experiments, we discussed similarities and differences in evolutionary dynamics concerning the Baldwin effect in a population of globally interacting agents and the one of locally interacting agents.

1 Introduction

The Baldwin effect [1] is known as one of the interactions between evolution and learning, which suggests that individual lifetime learning (phenotypic plasticity) can influence the course of evolution without the Lamarckian mechanism. Hinton and Nowlan clearly demonstrated this effect by a simple evolutionary simulation [2]. This had a large impact not only on biologists but also on computer scientists [3].

The Baldwin effect explains the interaction between evolution and learning by paying attention to balances between benefit and cost of learning through the following two steps [4]. In the first step, lifetime learning gives individual agents chances to change their phenotypes. If the learned traits are useful for agents and make their fitness increase, they will spread in the next population. The learning behavior acts as a benefit in this step. In the second step, if the environment is sufficiently sta-

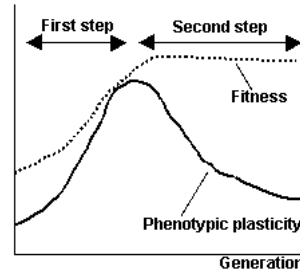


Figure 1: Two steps of the Baldwin Effect.

ble, the evolutionary path finds innate traits that can replace learned traits, because of the cost of learning. Through these steps, learning can accelerate the genetic acquisition of learned traits without the Lamarckian mechanism in general. Figure 1 roughly shows the concept of the Baldwin effect which consists of two steps described above.

Recently this effect has been discussed in various contexts. Mayley analysed the Baldwin effect in terms of its influence on the speed or rate of evolution by using Kauffman's NK fitness model [5, 6]. He showed the Hiding effect (learning slows the progress of an evolving population whilst producing fit phenotypes in contrast to the Baldwin effect) was observed, and whether the Baldwin effect or the Hiding effect dominates in the evolutionary scenario depends on the cost of learning and the ruggedness of the fitness landscape. Watson and Pollack adopted Hinton and Nowlan's model [2] by replacing lifetime plasticity with lifetime interactions between organisms [7]. They demonstrated how symbiotic relationships can guide the genetic make-up of organisms through two steps of the Baldwin effect. The Baldwin effect was also applied to the field of the hybrid genetic algorithm by Whitley, Gordon and Mathias [8]. They pointed out that the Baldwinian search strategy can sometimes be more effective than the Lamarckian search strategy in some popular function optimization problems. Although many studies have been conducted as mentioned above, most of them have discussed this effect on the assumption that environments are

Table 1: A payoff matrix of Prisoner’s Dilemma.

| Player \ Opponent | Cooperate | Defect |
|-------------------|------------|------------|
| | Cooperate | (R:3, R:3) |
| Defect | (T:5, S:0) | (P:1, P:1) |

(Player’s score, Opponent’s score)
 $T > R > P > S, 2R > T + S$

fixed and the optimal solution is unique. However, as we see in the real world, learning could be more effective and utilized in dynamic environments, because the flexibility of plasticity itself is advantageous to adapt ourselves to the changing world.

We have been investigating the Baldwin effect in the context of the evolution of game strategies [9, 10]. Our objective is to clarify the function and the mechanism of the Baldwin effect in dynamic environments focusing on balances between benefit and cost of learning that are caused by interactions among agents. We adopted the iterated Prisoner’s Dilemma (IPD) as dynamic environments, where there is no explicit optimal solution through generations and the fitness of agents depends mainly on interactions among them. Phenotypic plasticity, which can be modified by lifetime learning, has been introduced into strategies in our model, and we conducted computational experiments in preliminary experiments in which phenotypic plasticity is allowed to evolve. Subsequently, we introduced the spatial locality into our model and also conducted experiments with various scales of local interaction. Based on these experiments, we will discuss similarities and differences in evolutionary dynamics concerning the Baldwin effect between the population of globally interacting agents and the one of locally interacting agents.

2 The Model

2.1 Iterated Prisoner’s Dilemma Game

We have adopted the iterated Prisoner’s Dilemma (IPD) as a dynamic environment, which represents an elegant abstraction of the situations causing social dilemma. IPD game is carried out as follows:

1. Two players independently choose actions from Cooperate (C) or Defect (D) without knowing the other’s choice.
2. Each player gets the score according to the payoff matrix (Table 1). We term this procedure “round”.
3. Players play the round repeatedly, and compete for higher average scores.

In case of one round game, the payoff matrix makes defecting be the only dominant strategy re-

gardless of opponent’s action, and defect-defect action pair is the only Nash equilibrium. But this equilibrium is not Pareto optimal because the score of each player of cooperate-cooperate action pair is higher, which causes a dilemma. Furthermore, if the round is repeated in the game, cooperating with each other will turn out advantageous to both players in the long run.

2.2 Expression of Strategies for IPD

The strategies of agents are expressed by two types of genes: genes for representing strategy (GS) and genes for representing phenotypic plasticity (GP). GS describes deterministic strategies of IPD like Lindgren’s model [11], which defines next action according to the history of actions. GP expresses whether each corresponding bit of GS is plastic or not.

A strategy of memory m has an action history h_m which is a m -length binary string as follows:

$$h_m = (a_{m-1}, \dots, a_1, a_0)_2, \tag{1}$$

where a_0 is the opponent’s previous action (“0” represents defection and “1” represents cooperation), a_1 is the previous player’s action, a_2 is the opponent’s next to previous action, and so on.

GS for a strategy of memory m can be expressed by associating an action A_k (0 or 1) with each history k as follows:

$$GS = [A_0 A_1 \dots A_{n-1}] \quad (n = 2^m). \tag{2}$$

In GP , P_x specifies whether each phenotype of A_x is plastic (1) or not (0). Thus, GP can be expressed as follows:

$$GP = [P_0 P_1 \dots P_{n-1}]. \tag{3}$$

For example, the popular strategy “Tit for Tat” (cooperates in the first round, thereafter does whatever the opponent did in the previous round) can be described by memory 2 as $GS=[0101]$, $GP=[0000]$.

2.3 Meta-Pavlov Learning

A plastic phenotype can be changed by learning during game. We adopted a simple learning method termed “Meta-Pavlov”. Each agent changes plastic phenotypes according to the result of each round by referring to the Meta-Pavlov learning matrix (Table 2). It doesn’t express any strategy but expresses the way to change own strategy (phenotype) according to the result of the current round, though this matrix is the same as that of the Pavlov strategy [12]. The learning process is described as follows:

1. At the beginning of the game, each agent has the same phenotype as GS itself.
2. If the phenotype used in the last round is plastic, in other words, the bit of GP corresponding to the phenotype is 1, the phenotype

Table 2: The Meta-Pavlov learning matrix.

| Player \ Opponent | Cooperate | Defect |
|-------------------|-----------|--------|
| | Cooperate | C |
| Defect | D | C |

is changed to the corresponding value in the Meta-Pavlov learning matrix based on the result of the round.

3. The agent uses the new strategy specified by the modified phenotype from next round on.

Take a strategy of memory 2 expressed by $GS=[0001]$ and $GP=[0011]$, for example of learning (Figure 2). Each phenotype represents the next action corresponding to the history of the previous round, and the underlined phenotypes are plastic.

Let us suppose that the history of previous round was “C-C (player’s action: cooperation, opponent’s action: cooperation)” and the opponent defected at the present round. This strategy cooperates according to the phenotype and the result of the present round is “C-D” (the player gets Sucker’s payoff). The strategy changes own phenotype according to this failure based on the Meta-Pavlov learning matrix, because the phenotype applied at this round is plastic. The phenotype “C” corresponding to the history “C-C” is changed to “D” in this example. Therefore, this strategy always defects at the next round.

The values of GS that are plastic act merely as the initial values of phenotype. Thus we represent strategies by GS with plastic genes replaced by “x” (e.g. $GS=[1000]$, $GP=[1001] \Rightarrow [x00x]$).

2.4 Evolution

A population consists of N individuals that play the IPD game. All genes are set randomly in the initial population. Now, we assume the “global and non-spatial” interaction among agents: The round robin tournament is conducted between individuals with the strategies as described above. (The “local and spatial” interaction will be adopted in Section 4.) Performed action can be changed by

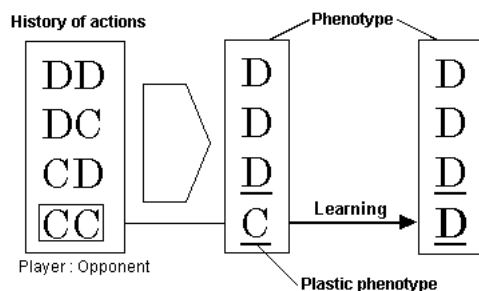


Figure 2: An example of Meta-Pavlov learning.

noise (mistake) with probability p_n . Each plastic phenotype is reset to the corresponding value of GS at the beginning of games. The game is played for several rounds. We shall assume that there is a constant probability p_d (*discount parameter*) for another round. The tournament is “ecological”: The total score of each agent is regarded as a fitness value, new population is generated by the “roulette wheel selection” according to the scores, and mutation is performed on a bit-by-bit basis with probability p_m .

Average scores during the first 20 IPD games between new pairs are stored, and will be used as the results of the games instead of repeating games actually, so as to reduce the amount of computation. Stored scores are cleared and computed again by playing games every 500 generation. (Though this method could reduce the influence of p_n or p_d , it has turned out not to affect the essential character of evolutionary phenomena.)

3 Experiments on Globally Interacting Agents

Strategies of memory 2 were investigated in the preliminary experiments. We conducted an evolutionary experiment for 2000 generations using the following parameters: $N = 1000$, $p_m = 1/1500$, $p_n = 1/25$ and $p_d = 99/100$.

The results are shown in Figure 3 and 4. In each figure, the horizontal axis represents the generations. The vertical axis represents the distribution of strategies and also represents both “plasticity of population” (in black line) which is the ratio of “1” in all GP s and the average score (in white line). Plasticity of population is supposed to correspond to the “phenotypic plasticity” in Figure 1. The average score represents the degree of cooperation in the population, and it takes 3.0 as the maximum value when all rounds are “C-C”.

The evolutionary phenomena observed in a typical experiment are summarized as follows. Defect-oriented strategies ($[0000]$, $[000x]$, and so on) spread and made the average score decrease until about 60th generation, because these strategies can’t cooperate with each other. Simultaneously, partially plastic strategies ($[0x0x]$, $[00xx]$, and so on) occupied the population. Next, around the 250th generation, more plastic strategies ($[xxxx]$, $[x0xx]$, and so on) established cooperative relationships quickly, which made the plasticity and average fitness increase sharply. This transition is regarded as the first step of the Baldwin effect.

Subsequently, the plasticity of the population decreased and then converged to 0.5 while keeping the average score high. Finally, the strategy $[x00x]$ occupied most of the population. The reason seems to be that the strategy has the necessary and sufficient amount of plasticity to maintain cooperative relationships and prevent other strategies from in-

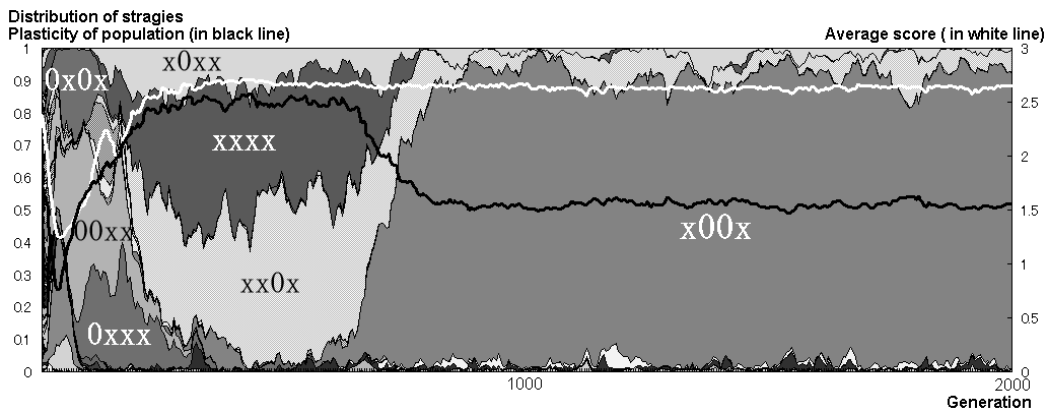


Figure 3: The experimental result (2000 generations).

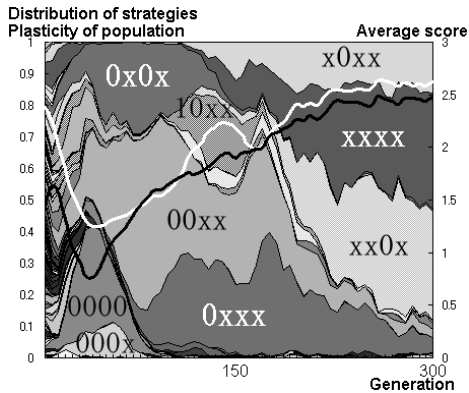


Figure 4: The experimental result (300 generations).

vading the population. This transition is regarded as the second step of the Baldwin effect.

The population converged to the strategy $[x00x]$ in almost all experiments, and the evolutionary phenomena described above were observed in 70% of experiments. Further analysis has shown remarkable properties of the strategy $[x00x]$, which are summarized as follows:

- The strategy $[x00x]$ satisfies the ESS (Evolutionarily Stable Strategy) condition in all 256 strategies of memory 2.
- The state transition analysis shows that the minimal actions for recovery of mutual cooperation from an accidental defection (D^*) by a duration of mutual cooperation are “C- D^* (or D^* -C), D-D, C-C, D-D, C-C” when one strategy plays against $[x00x]$. This “fence-mending” protocol is performed exactly when $[x00x]$ plays against $[x00x]$ itself.
- The left “x” (describes the plasticity of the action immediately after “D-D”) is effective especially when $[x00x]$ plays against defect-oriented

strategies, the right “x” (describes the plasticity of the action immediately after “C-C”) is effective especially when $[x00x]$ plays against cooperate-oriented strategies, and the minimal protocol described above is realized by utilizing both of plastic genes.

4 Experiments on Locally Interacting Agents

We had adopted the non-spatial model in which one agent plays games against all the other agents in the preliminary experiment. However, the strength of spatial locality must have some kind of effect on the course of evolution, because the fitness of each agent is evaluated only by interactions among agents in our model.

Many studies have focused on the effects of spatial locality in the interactions between players, without taking interest in the evolution of the plasticity. In general, cooperative strategies can easily occupy the population in the spatial IPD model compared to the non-spatial model, because cooperative strategies can make a cluster of own strategies while defect-oriented strategies can not. Actually, Grim reported that the a strategy, which is twice as generous as the “GTFT” (cooperates on the first round, cooperates with probability 1/3 when the opponent defected on the previous round, otherwise does whatever its opponent did) [13] turned out to be optimal with just a two-dimensional spatialization of the stochastic IPD model [14].

Here, we expanded our model in two-dimensional space where each agent plays games only with its neighbors so as to investigate the local interactions in our model. We modified the existing model as follows:

- There are troidal $W \times W$ lattice sites and each site contains a single strategy which plays the IPD game. All genes are set randomly in the initial population.

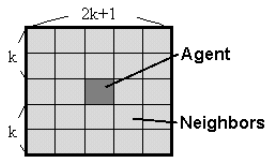


Figure 5: The nearest $(2k + 1)^2 - 1$ neighbors ($k = 2$).

- Each agent plays iterated games against $(2k + 1)^2 - 1$ neighbors as defined in Figure 5.
- The agent that occupies each site in the next generation is selected from nearest $(2k + 1)^2$ agents (the agent in each site and its neighbors) by the “roulette wheel selection” according to the scores.

The parameter k represents the scale of interaction because it determines the number of neighbors for each agent. For example, when $k = 1$, each agent plays against only 8 neighbors; when $k = (W - 1)/2$, the model has no spatial locality and is equivalent to the non-spatial model in the previous section because each agent plays all the other agents. Lindgren also investigated the diversity of strategies with various payoff matrixes of IPD game by using a spatial IPD model [15]. Our model is different from his model because the deterministic reproduction and the fixed number of neighbors were adopted in his model. By altering the scale of local interaction, we analyze the influence that local interactions have on the course of evolution and the plasticity of population.

We conducted the evolutionary experiment using the following parameters through 2000 generations: $W = 33$ ($N = 1089$), $k = 1, 2, 3, 5, 10$, and 16, and all the other parameters are the same as adopted in Section 3. Figure 6 shows the average score and Figure 7 also shows the transitions of the plasticity of population with various values of k . Each experimental result is the average of 100 trials.

The results show that the scale of interaction affects the emergence of the Baldwin effect in our model. When k was relatively low (such as $k = 1$ or 2), the invasion of defect-oriented strategies in the early period (until about 60th generation) was observed but the decline of the average score was smaller and the minimum was not as deep as for larger values of k . Subsequently the plasticity of population slowly increased, and the population easily established cooperative relationships then finally converged to the strategy [x00x].

As the value of k increased, the plasticity rose to higher peaks and the Baldwin effect was more clearly observed, although there were several exceptions. This is because as the scale of interaction

becomes larger, the influence of the spatial locality becomes smaller; cooperative strategies become difficult to spread. Therefore sufficient plasticity was required for a shift from defect-oriented to cooperative populations. This is the reason why the benefit of plastic behavior emerged when k was relatively high. Contrarily, the plastic behavior was not required to establish the cooperative relationships when the value of k was relatively low.

Also, we found another role of the spatial locality. When k was relatively high ($k \geq 3$), the average fitness rose to about 2.6 more slowly in the second step of the Baldwin effect. This slight increase is due to the exceptional case of evolutionary scenario which is different from that described in the previous section. That is, defect-oriented strategies temporarily occupied the population, then made the population unstable and the average score decreased without spatial locality. This result shows that the cost of learning was also emerged in the globally interacting environment.

Kauffman studied the dynamics of coevolving systems using the NKC model [5]. Briefly speaking, he pointed out that as the parameter C (the number of traits which affect the fitness of the other species) increases the environment becomes more chaotic. The scale of interaction k in our model approximately corresponds to the parameter C in his model. These results partly reflect his claim and suggest that the phenotypic plasticity could make such a complex system well-ordered.

5 Conclusion

We have discussed how learning can affect the course of evolution in dynamic environments based on the results of computational experiments on the evolution of game strategies. In the non-spatial model, when we introduced the Meta-Pavlov learning to strategies as phenotypic plasticity, population evolved to be cooperative and stable through two steps of the Baldwin effect.

We also investigated how local interactions among agents affect the benefit and the cost of learning using the two-dimensional version of our model. Experiments with various scales of interaction have shown two aspects concerning the relation between spatial locality and phenotypic plasticity: First, the Baldwin effect is sensitive to the scale of local interaction; Second, as agents interact more globally, the evolutionary scenario becomes unstable. The reason is that the benefit and the cost of learning depend on the scale of interaction, although they are not explicitly embedded in our model.

References

- [1] J. M. Baldwin. A New Factor in Evolution. *American Naturalist*, Vol. 30, pp. 441–451, 1896.

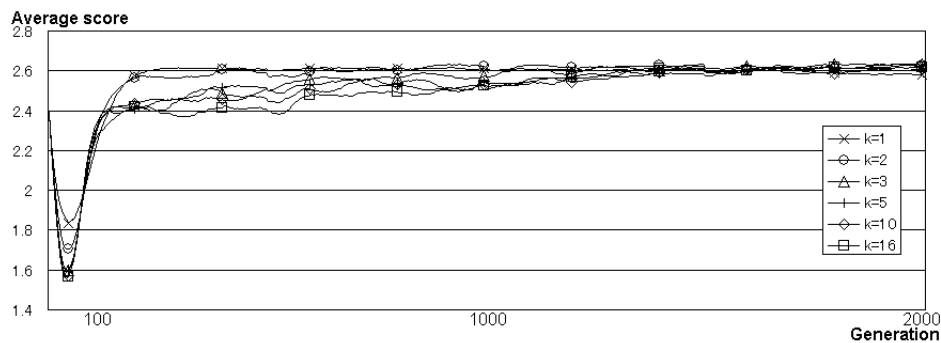


Figure 6: Transitions of the average score (2000 generations).

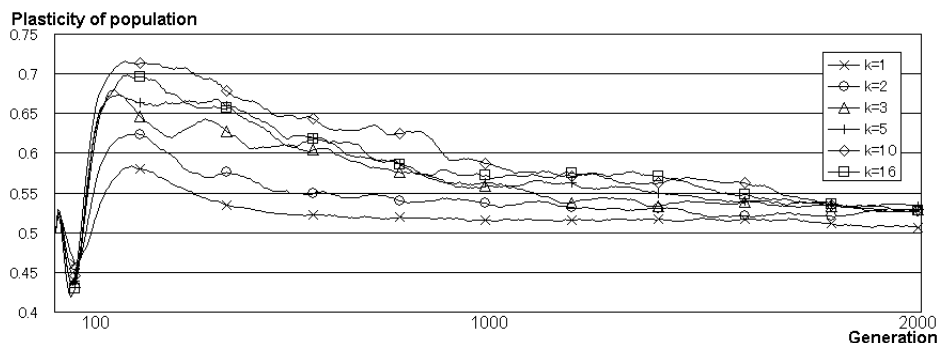


Figure 7: Transitions of the plasticity of population (2000 generations).

- [2] G. E. Hinton and S. J. Nowlan. How Learning Can Guide Evolution. *Complex Systems*, Vol. 1, pp. 495–502, 1987.
- [3] T. Arita. *Artificial Life: A Constructive Approach to the Origin/Evolution of Life, Society, and Language*. Science Press, 2000 (in Japanese).
- [4] P. Turney, D. Whitley and R. W. Anderson. Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect. *Evolutionary Computation*, Vol. 4, No. 3, pp. 4–8, 1996.
- [5] S. A. Kauffman. *The Origins of Order: Self Organization and Selection in Evolution*. Oxford University Press, 1993.
- [6] G. Mayley. Guiding or Hiding: Explorations into the Effects of Learning on the Rate of Evolution. *Proceedings of Fourth European Conference on Artificial Life*, pp. 135–144, 1997.
- [7] R. A. Watson and J. B. Pollack. How symbiosis Can Guide Evolution. *Proceedings of Fifth European Conference on Artificial Life*, pp. 29–38, 1999.
- [8] D. Whitley, V. S. Gordon and K. Mathias. Lamarckian Evolution, The Baldwin Effect and Function Optimization. *Parallel Problem Solving from Nature III*, pp. 6–15, 1994.
- [9] R. Suzuki and T. Arita. How Learning Can Affect the Course of Evolution in Dynamic Environments. *Proceedings of Fifth International Symposium on Artificial Life and Robotics*, pp. 260–263, 2000.
- [10] T. Arita and R. Suzuki. Interactions between Learning and Evolution: The Outstanding Strategy Generated by the Baldwin Effect. *Proceedings of Artificial Life VII* pp. 196–205, 2000.
- [11] K. Lindgren. Evolutionary Phenomena in Simple Dynamics. *Proceedings of Artificial Life II*, pp. 295–312, 1991.
- [12] M. A. Nowak and K. Sigmund. A Strategy of Win-Stay, Lose-Shift that Outperforms Tit-for-Tat in the Prisoner’s Dilemma Game. *Nature*, Vol. 364, pp. 56–58, 1993.
- [13] M. A. Nowak and K. Sigmund. Tit for Tat in Heterogeneous Populations. *Nature*, Vol. 355, pp. 250–252, 1992.
- [14] P. Grim. Spatialization and Greater Generosity in the Stochastic Prisoner’s Dilemma. *BioSystems*, Vol 37, pp. 3–17, 1996.
- [15] K. Lindgren and M. G. Nordahl. Evolutionary Dynamics of Spatial Games. *Physica D*, Vol. 75, pp. 292–309, 1994.