Interactions between Learning and Evolution: The Outstanding Strategy Generated by the Baldwin Effect

Takaya Arita and **Reiji Suzuki**
Graduate School of Human Informatics, Nagoya University
Chikusa-ku, Nagoya 464-8601, JAPAN

Abstract

The Baldwin effect is known as interactions between learning and evolution, which suggests that individual lifetime learning can influence the course of evolution without the Lamarckian mechanism. Our concern is to consider the Baldwin effect in dynamic environments, especially when there is no explicit optimal solution through generations and it depends only on interactions among agents. We adopted the iterated Prisoner's Dilemma as a dynamic environment, introduced phenotypic plasticity into strategies, and conducted the computational experiments, in which phenotypic plasticity is allowed to evolve. The Baldwin effect was observed in the experiments as follows: First, strategies with enough plasticity spread, which caused a shift from defective population to cooperative population. Second, these strategies were replaced by a strategy with a modest amount of plasticity generated by interactions between learning and evolution. By making three kinds of analysis, we have shown that this strategy provides the outstanding performance. Further experiments towards open-ended evolution have also been conducted so as to generalize our results.

Introduction

Baldwin proposed 100 years ago that individual lifetime learning (phenotypic plasticity) can influence the course of evolution without the Lamarckian mechanism (Baldwin, 1896). This "Baldwin effect" explains the interactions between learning and evolution by paying attention to balances between benefit and cost of learning. The Baldwin effect consists of the following two steps (Turney, Whitley and Anderson, 1996). 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. This step means the synergy between learning and evolution. In the second step, if the environment is sufficiently stable, the evolutionary path finds innate traits that can replace learned traits, because of the cost of learning. This step is known as *genetic assimilation*. 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.

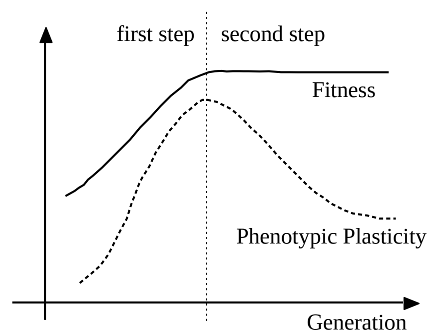

Figure 1: Two steps of the Baldwin effect.

Hinton and Nowlan constructed the first computational model of the Baldwin effect and conducted an evolutionary simulation (Hinton and Nowlan, 1987). Their pioneering work made the Baldwin effect come to the attention of the computer scientists, and many computational approaches concerning the Baldwin effect have been conducted since then (Arita, 2000). For example, Ackley and Littman successfully showed that learning and evolution together were more successful than either alone in producing adaptive populations in an artificial environment that survived to the end of their simulation (Ackley and Littman, 1991). Also, Bull recently examined the performance of the Baldwin effect under varying rates and amounts of learning using a version of the NK fitness landscapes (Bull, 1999).

Most of them including Hinton and Nowlan's work have assumed that environments are fixed and the optimal solution is unique, and have investigated the first step (synergy between learning and evolution). 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. Therefore, it is essential to examine how learning can affect the course of evolution in dynamic environments (Suzuki and Arita, 2000).

Our objective is to have a valuable insights into interactions between learning and evolution, especially into the Baldwin effect, by focusing on balances between benefit and cost of learning in dynamic environments: whether the Baldwin effect is observed or not, how it works, and what it brings after all in dynamic environments.

As one of the few studies that looks at both the benefits and costs of learning (in static environments), Menczer and Belew showed that interactions between learning and evolution are not beneficial if the task that learning is trying to optimize is not correlated with the task that evolution is working on (Menczer and Belew, 1991). Also, Mayley explored two criteria for the second step of the Baldwin effect by using NK fitness landscapes (Mayley, 1997). He concluded that two conditions, high relative evolutionary cost of learning and the existence of a neighborhood correlation relationship between genotypic space and phenotypic space, are the necessary conditions for the second step to occur.

In general, dynamic environments can be divided typically into the following two types: the environments in which the optimal solution is changed as the environment changes, and the ones in which each individual's fitness is decided by interactions with others. As the former type of environments, Anderson quantitatively analyzed how learning affects evolutionary process in the dynamic environment whose optimal solution changed through generations by incorporating the effects of learning into traditional quantitative genetics models (Anderson, 1995). It was shown that in changing environments, learning eases the process of genetic change in the population, while in fixed environments the individual advantage of learning is transient. Also, Sasaki and Tokoro studied the relationship between learning and evolution using a simple model where individuals learned to distinguish poison and food by modifying the connective weights of neural network (Sasaki and Tokoro, 1999). They have shown that the Darwinian mechanism is more stable than the Lamarckian mechanism while maintaining adaptability. Both studies emphasized the importance of learning in dynamic environments.

We adopted the iterated Prisoner's Dilemma (IPD) as the latter type of environments, where there is no explicit optimal solution through generations and fitness of individuals 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 the computational experiments in which phenotypic plasticity is allowed to evolve.

Rest of the paper is organized as follows. Section 2 describes a model for investigating the interactions between learning and evolution by evolving the strategies for the IPD. The results of evolutionary experiments based on this model are described in Section 3. In Section 4, we analyze the strategy generated by the Baldwin effect in these experiments by three methods (ESS condition, state transition analysis and qualitative analysis). Section 5 describes the extended experiments towards open-ended evolution in order to generalize the results in the previous sections. Section 6 summarizes the paper.

Model

Expression of Strategies for the Prisoner's Dilemma

We have adopted the iterated Prisoner's Dilemma (IPD) game 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 game repeatedly, retaining access at each round to the results of all previous rounds, and compete for higher average scores.

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

player \ opponent	cooperate	defect
cooperate	(R:3, R:3)	(S:0, T:5)
defect	(T:5, S:0)	(P:1, P:1)

(player's score, opponent's score)
$T > R > P > S, 2R > T + S$

In case of one round game, the payoff matrix makes defecting be the only dominant strategy regardless 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 is higher when both of the players cooperate, which causes a dilemma. Furthermore, if the same couple play repeatedly, this allows each player to return the co-player's help or punish co-player's defection, and therefore cooperating each other can be advantageous to both of them in the long run (Axelrod, 1984).

The strategies of agents are expressed by two types of genes: genes for representing strategies (GS) and genes for representing phenotypic plasticity (GP). GS describes deterministic strategies for IPD by the method adopted in Lindgren's model (Lindgren, 1991), 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:

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

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:

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

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

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

For example, the popular strategy "Tit-for-Tat" (cooperates on the first round, does whatever its opponent did on the previous round) (Axelrod, 1984) can be described by memory 2 as GS=[0101] and GP=[0000].

Meta-Pavolv 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 which is famous because it was shown that it outperforms the popular strategy "Tit-for-Tat" (Nowak and Sigmund, 1993).

Table 2: The Meta-Pavlov learning matrix.

player \ opponent	cooperate	defect
cooperate	C	D
defect	D	C

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 was plastic, in other words, the bit of GP corresponding to the phenotype is 1, the phenotype is changed to the corresponding value in the Meta-Pavlov learning matrix based on the result of the last round.

3) The new strategy specified by the modified phenotype will be used by the player 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.

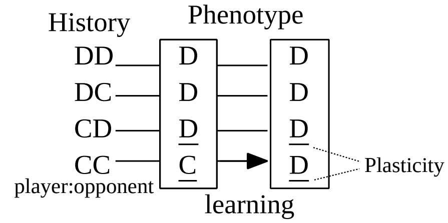

Figure 2: An example of Meta-Pavlov learning.

Let us suppose that the action pair of the previous round was "CC (player's action: cooperation, opponent's action: cooperation)" and the opponent defects at the present round. This strategy cooperates according to the phenotype and the result of the current round is "CD" (*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 "CC" is changed to "D" in this example. Therefore, this strategy chooses defection when it has the history "CC" at the next time. Meta-Pavlov learning is intuitive and natural in the sense that it is a simple realization of reinforcement learning.

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] and GP=[1001] $\rightarrow$ [x00x]).

Evolution

We shall consider a population of N individuals interacting according to the IPD. All genes are set randomly in the initial population. The round robin tournament is conducted between individuals with the strategies which are expressed in the above described way. Performed action can be changed by 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 pair 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 doing games every 500 generation.

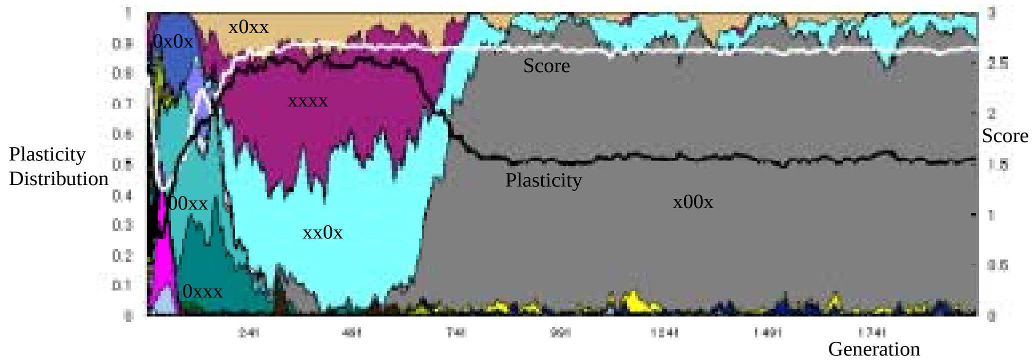

Figure 3: The experimental result (2000 generations).

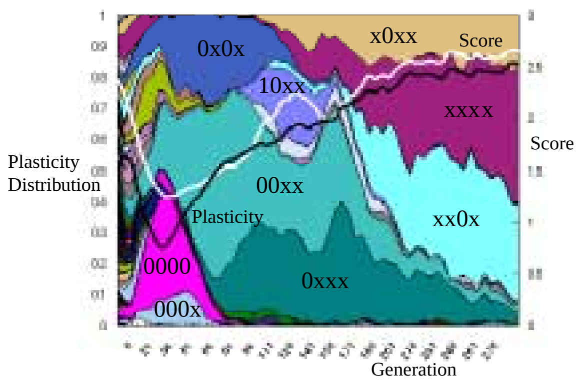

Figure 4: The experimental result (300 generations).

Evolutionary Experiments

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

The evolution of population for the first 2000 generations is shown in Figure 3 and that for the first 300 generations is shown in Figure 4. In each figure, the horizontal axis represents the generations. The vertical axis represents the distribution of strategies, and at the same time, it also represents both "plasticity of population" (in black line) and the average score (in white line). Plasticity of population is the ratio of "1" in all genes of GPs, and it corresponds 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 "CC".

The evolutionary phenomena that were observed in experiments are summarized as follows. Defective strategies ([0000], [000x] and so on) spread and made the average score decrease until about 60th generation, because these strategies can't cooperate each other. Simultaneously, partially plastic strategies ([0x0x], [00xx] and so on) occupied most of 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 population decreased and then converged to 0.5 while keeping the average score high. Finally, the strategy [x00x] occupied 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 invading in the population. This transition is regarded as the second step of the Baldwin effect.

The evolutionary phenomena described above was observed in about 70% of the experiments, and the population converged to the strategy [x00x] in *all* experiments we conducted. Further analysis on this strategy will be conducted in the next Section. Another series of experiments has shown overall that the higher the mutation rate becomes, the faster the strategies tend to evolve. It has been also shown that the higher the noise probability becomes, the more All-D type strategies are selected, and the less the system becomes stable.

Figure 5 made us grasp the clear image of the evolutionary behavior of the system in the experiments. This figure shows the evolutionary trajectory of ten experiments drawn in the space of score and plasticity. We see the evolutionary process consists of 3 modes. The score decreases without increase of plasticity during an initial stage. The cause of this decrease is that defect-oriented strategies (e.g. [0000][000x]) spread in the initial randomly-created population. The score decreases nearly to 1.0 which is the score in the case of defect-defect action pair. When the score reaches this value, a "mode transition" happens and the first step of the Baldwin effect starts. In this stage, phenotypic plasticity gives chances to be adaptive. Therefore, score is correlated with plasticity, and approaches nearly 3.0, that is the

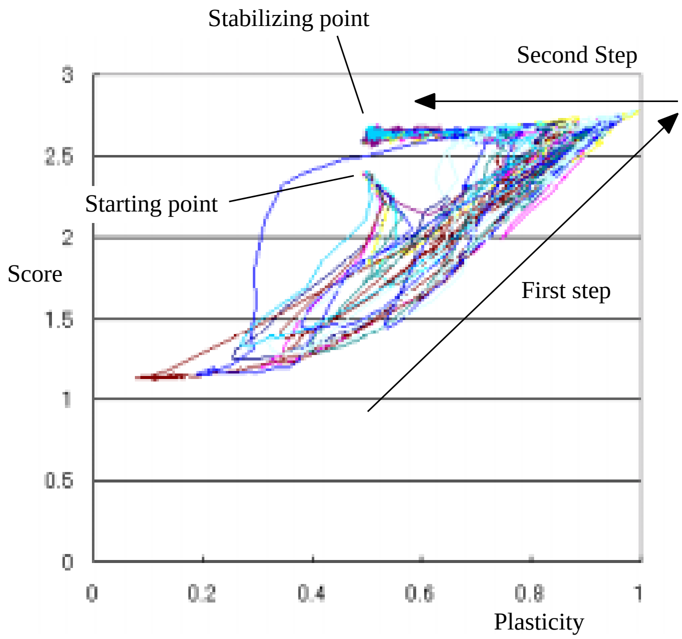

Figure 5: Two steps of the Baldwin effect.

score in the case of cooperate-cooperate action. Strategies with enough plasticity (e.g. [xxxx][x0xx][xx0x]) occupy at the end of this stage. Then, another mode transition happens suddenly, and plasticity decreases gradually while keeping score high. The plasticity decreases monotonously, and after all, the population always converged to be homogenous that is occupied with the strategy [x00x]. As is apparent from this figure, there were exceptions to which above description doesn't apply, however, it has been shown that the system always stabilized with [x00x] after all.

Analysis of Meta-Pavlov [x00x]

ESS Condition

An ESS (Evolutionary Stable Strategy) is a strategy such that, if all the members of a population adopt it, no mutant strategy can invade (Maynard Smith, 1987). The necessary and sufficient condition for a strategy "a" to be ESS is:

$$E(a,a) > E(b,a) \quad \forall b, \tag{4}$$

or

$$E(a,a) = E(b,a) \quad and \quad E(a,b) > E(b,b) \quad \forall b, \tag{5}$$

where $E(a,b)$ is the score of strategy "a" when strategy "a" plays against strategy "b".

We conducted the iterated games between [x00x] (GS=[0000], GP=[1001]) and all 256 strategies with memory 2, and computed the average scores of them, so as to examine whether it satisfied the ESS condition or not. The noise probability (p_n) was 1/25 and the discount parameter (p_d) was 99/100. The results are shown in Figure 6. The horizontal axis represents all strategies by interpreting the genotypic expression [$GSGP$] as an 8 bit binary number x (e.g. GS=[0000], GP=[1001] $\rightarrow$ 00001001_2=9). The vertical axis represents the relative scores of the strategy x, that is,

$$E([\text{x00x}],[\text{x00x}]) - E(x,[\text{x00x}]). \tag{6}$$

This graph shows that this value is always positive. Therefore, [x00x] is an ESS in the population of memory 2 strategies.

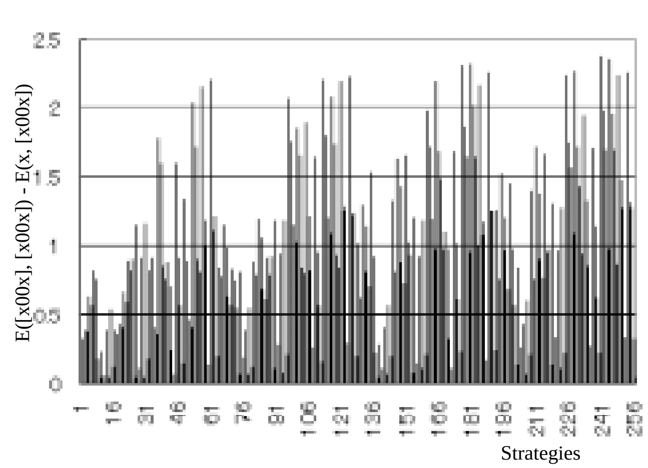

Figure 6: Relative scores of all strategies of memory 2 against [x00x].

State Transition Analysis

Figure 7 shows a state transition diagram of the Meta-Pavlov [x00x] strategy. Each state is represented by a box, in which the actions in the current round are described: the opponent's action on top and the [x00x] 's action on bottom (0: defect, 1: cooperate). The current values of plastic genes also discriminate the states, and they are described in the lower right corner (e.g. left "x"=0 and right "x"=1 $\rightarrow$ 01). Two arrows issue forth from each state, depending on whether the opponent plays C or D at the next round. Described actions of [x00x] in the destination box are identical, and it will be the next action of [x00x]. For example, the stabilized state of the game between [x00x] and All-D is expressed by a loop ("cycle 2" in this figure), which means that the game generates the periodic action pairs. The boxes without inputted arrows can be reached by noise.

Duration of the state "A" means that mutually cooperative relationship has been established. It is a remarkable point that if this relationship is abolished by the opponent's defective action, a bit of *protocol* (cycle "1" in this figure) is needed to restore the damaged relationship as follows:

```
Opponent: .. 1  1  0  0  1  0  1  1  1  ..
[x00x]:   .. 1  1  1  0  1  0  1  1  1  ..
       Mutual C |noise|fence-  | Mutual C
                      mending
```

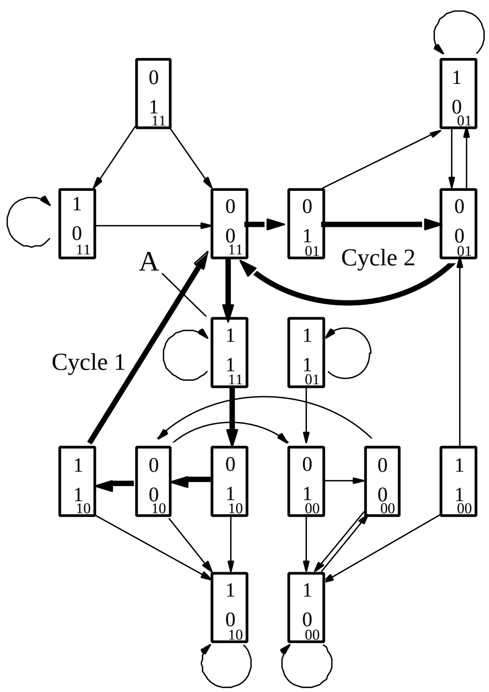

Figure 7: A state transition diagram of the Meta-Pavlov [x00x].

This minimal fence-mending is done exactly when an accidental opponent's defect by noise occurs after mutually cooperative relationship has been established in the game between [x00x] and itself. This property of [x00x] seems to play an important part in recovery from the broken relationship, and to make the strategy be an ESS.

Qualitative Analysis on Phenotypic Plasticity

Many researchers in evolutionary computation or related fields have focused exclusively on the benefits on the phenotypic plasticity. Phenotypic plasticity enables the individuals to explore neighboring regions of phenotype space. The fitness of an individual is determined approximately by the maximum fitness in its local region. Therefore, if the genotype and the phenotype are correlated, plasticity has the effect of smoothing the fitness landscape, and makes it easier for evolution to climb to peaks in the landscape. This is the first step of the Baldwin effect.

However, there is the second step, because plasticity can be costly for an individual. Learning requires energy and time, for example, and sometimes brings about dangerous mistakes. In our computational experiments, the costs of learning are not explicitly embed in the system. The costs of learning are implicitly expressed by the behavior that are caused typically by noise. For example, when a noise happens to a game between [x00x] and [xxxx], plastic properties make the [xxxx] strategy play more C than [x00x] while they restore the damaged relationship, which generates [xxxx]'s loss. The optimum balance between plasticity and rigidity depends on the performance of the learning algorithm. In this context, Meta-Pavlov learning algorithm gets along extremely well with [x00x], as will be shown in the extended experiments.

Here, we investigate why these two plastic genes in [x00x] remained in the second step of the Baldwin effect, that is, the significance of the two plastic genes. While the functions of these two genes are of course depend on the interactions among all genes, simple explanation could be possible based on the results of our qualitative analysis as follows:

- The left "x" (which describes the plasticity of the action immediately after D-D) is effective especially when [x00x] plays against defect-oriented strategies. For example, when [x00x] plays against All-D, [x00x] gets the *Sucker's* payoff once every three rounds (cycle 2 in Figure 7), and gets only 0.67 on average, caused by the plasticity of the left "x". However, All-D gets about 2.33, which supports the ESS property of [x00x] because [x00x] gets about 2.6 when it plays against itself, as follows:

  ```
  [0000]:  .. 000000000000 ..  Average 2.33
  [x00x]:  .. 010010010010 ..  Average 0.67
  ```

  In contrast, for example, the game between the Pavlov strategy (Nowak and Sigmund, 1994) and All-D is as follows:

  ```
  [0000]:  .. 000000000000 ..  Average 3
  [1001]:  .. 010101010101 ..  Average 0.5
  ```

- The right "x" (which describes the plasticity of the action immediately after C-C) is effective especially when [x00x] plays against cooperate-oriented strategies. [x00x] can defect flexibly by taking advantage of the opponent's accidental defect by noise. The right "x" becomes 0, when C-C relationship is abolished by the opponent's defect as shown in Figure 7. Therefore, [x00x] exploits relatively cooperate-oriented strategies. Followings are the rounds between [x001] and [x00x]. The first 0 of [x001] represents an accidental defect by noise. Average scores are calculated only during the oscillation.

  ```
  [x001]:  .. 1110011011011 ..  Average 1.33
  [x00x]:  .. 1111010010010 ..  Average 3
  ```

On the other hand, for example, the game between the Pavlov strategy and [x001] is as follows:

```
[x001]:  .. 1110000000000 ..  Average 3
[1001]:  .. 1111010101010 ..  Average 0.5
```

These two properties of [x00x] are quite effective on the premise that it establishes strong relationship with itself. Actually, minimal fence-mending is realized by utilizing these two plastic genes (two times of learning each gene) which is represented by the "cycle 1" in Figure 7.

Extended Experiments towards Open-ended Evolution

Evolution of Learning Algorithms

We have adopted the Meta-Pavlov learning method as an algorithm for modifying strategies by changing plastic phenotype so far. Here, we weaken this constraint, and shall focus on the evolution of not only strategies but also learning algorithms by defining the third type of genes.

In the experiments described in this section, each individual has genes for defining a learning method (GL), which decides how to modify the phenotype representing its strategies. GL is a four-length binary string composed of the elements of learning matrix such like Table 2. The order of elements in the string is [(DD) (DC) (CD) (CC)]. For example, the Meta-Pavlov learning method described in the previous sections is expressed by [1001]. It could be said that the learning methods (GL) and the strategies (GS and GP) co-evolve, because the performance of learning methods depends on the strategies to which they will be applied.

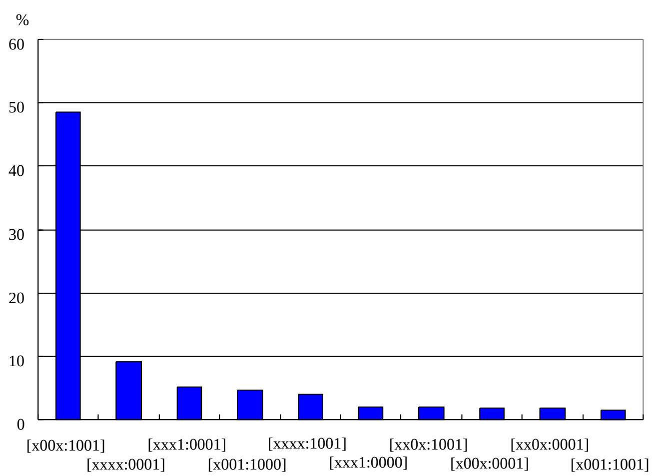

Figure 10: Average occupation of strategies.

Experiments were conducted under the same conditions as those in the previous experiments except for GL. Initial population had 100 kinds of combinations of randomly generated GS, GP and GL, and each kind had ten identical individuals. Typical results are shown in Figure 8 and Figure 9. Each area in these figures expresses a (strategy, learning method) pair. For example, "x00x:1001" means the [x00x] strategy with the learning method [1001] (Meta-Pavlov). It is shown that Meta-Pavlov [x00x] and [x001:1000] occupied the populations and established a stable state in Figure 8 and Figure 9 respectively.

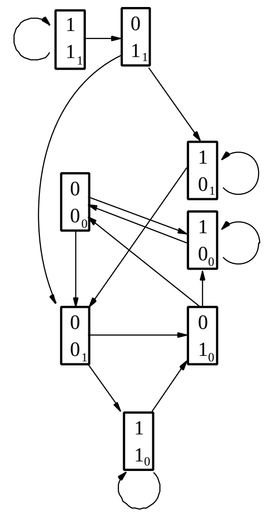

Figure 11: A state transition diagram of [x001:1000].

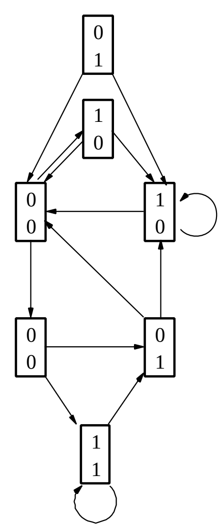

Figure 12: A state transition diagram of Prudent-Pavlov.

Figure 10 shows the average occupation of top ten (strategy, learning method) pairs in the 4000th generation over 60 trials. It is shown that Meta-Pavlov [x00x] occupied nearly half of the population in the 4000th generation on average. Meta-Pavlov [x00x] occupied the

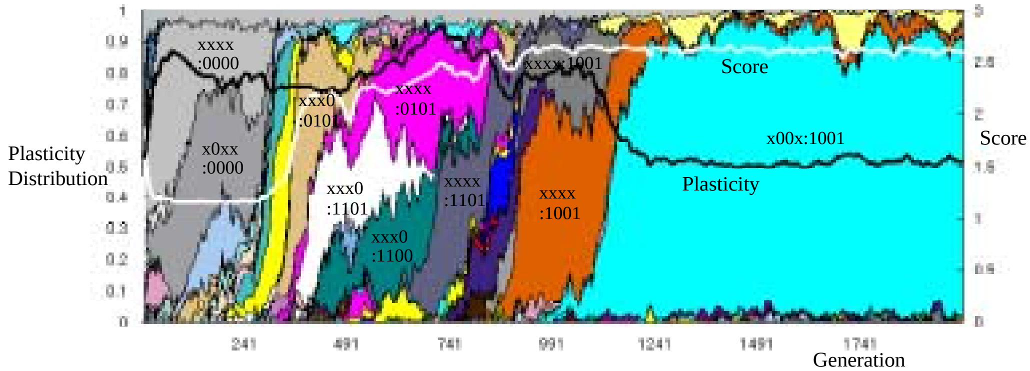

Figure 8: Evolution of learning algorithms and strategies (Case 1).

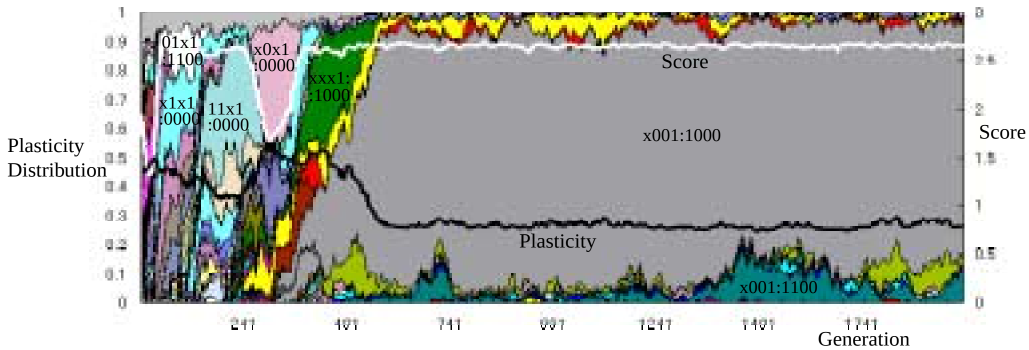

Figure 9: Evolution of learning algorithms and strategies (Case 2).

population and established a stable state (as shown in Figure 8) in 29 trials, [x001:1000] which is at the 4th in Figure 10 did so (as shown in Figure 9) in 3 trials, and no pairs occupied the population and established a stable state in the rest of trials. It follows from these facts that all but these two strategies in Figure 10 are invaded by mutants, though they can invade the population in certain conditions.

A state transition diagram of [x001:1000] is shown in Figure 11. We have found that this strategy has essentially the same property as that of "Prudent-Pavlov", whose state transition diagram is shown in Figure 12, though [x001:1000] has additional transient nodes, and there are subbtle differences in expression of states and state transitions. Prudent-Pavlov can be interpreted as a sophisticated offspring of Pavlov (Boerljst, Nowak and Sigmund, 1997). Prudent-Pavlov follows in most cases the Pavlov strategy. However, after any defection it will only resume cooperation after two rounds of mutual defection. They are remarkable facts that in our experiments a derivative of such a sophisticated *human-made* strategy was generated automatically, and that the Meta-Pavlov [x00x] outperformed the other strategies including this strategy.

Evolution without limitation of memory length

We have conducted further experiments towards open-ended evolution. Two types of mutation, gene duplication and split mutation, were additionally adopted, which allows strategies to become complex or simple without restrictions. The gene duplication attaches a copy of the genome itself (e.g., [1101] → [11011101]). The split mutation randomly removes the first or second half of the genome (e.g., [1101] → [11] or [01]). Each mutation is operated on GS and GP at the same time. In this series of experiments, we adopted Meta-Pavlov learning without allowing the learning mechanisms to evolve for convenience of the analysis.

Initial population was composed of strategies of memory 1, each of which has randomly generated GS and GP which was set to [00] (no plasticity). The results are shown in Figure 13. In most trials, during the first hundreds of generations, the system oscillated ([01] → [11] → [10] → [00]) in the same manner as in the Lindgren's experiments (Lindgren, 1991). At the end of the period of oscillation, a group of memory 2 strategies was growing, and took over the population. After that, there were two major evolutionary pathways, both of which happened

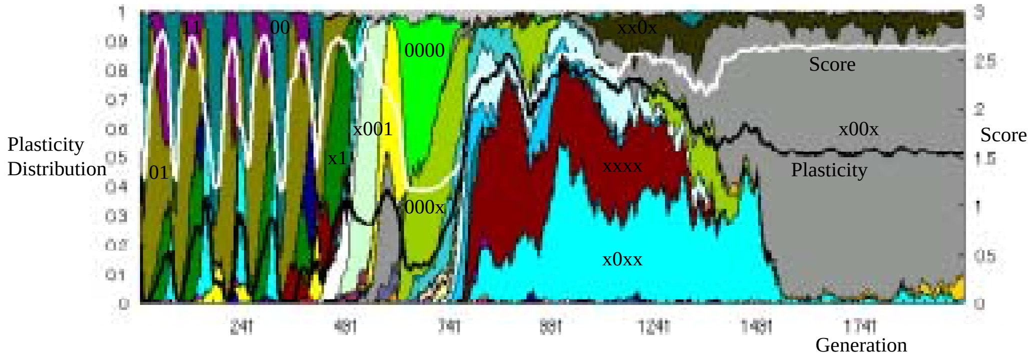

Figure 13: Evolution without limitation of memory length.

with nearly equal probabilities:

1) Strategies evolved showing the Baldwin effect as described in the previous sections. Later on, the system stabilized with [x00x] typically near the 1500th generation.

2) [x00x] entered the scene quickly, took over the generation, and the system stabilized with it.

It has been shown that which course the evolution takes depends on the state of the population while memory 2 strategies is growing. If the population is taken over by defect-oriented strategies before cooperate-oriented strategies emerge, the evolution tends to take the course 1). On the other hand, if the population is taken over by cooperate-oriented strategies without emergence of defect-oriented strategies of memory 2, then the evolution tends to take the course 2).

In most cases we observed, the system got stuck in the evolutionary stable state through either of the courses, though in rare cases the system didn't stabilize with [x00x] but stabilized with some mixture of various strategies of more than 2-length memory. The reason why strategies of more than 2-length memory rarely evolved is considered to relate to the mutation of learning mechanisms. The point here is that gene duplication changes the phenotype corresponding to the plastic genes because learning happens independently at two different points if a plastic gene is duplicated. Therefore, the evolution of phenotype could be discontinuous when gene duplication happens.

Conclusion

The Baldwin effect has not always been well received by biologists, partly because they have suspected it of being Lamarckist, and partly because it was not obvious it would work (Maynard Smith 1996). Our results of the experiments inspire us to image realistically how learning can affect the course of evolution in dynamic environments. It is a remarkable fact that a drastic mode transition happens at the edge between the first step and second step of the Baldwin effect in the environments where the optimal solution is dynamically changed depending on the interactions between individuals as is clearly shown in Figure 5.

Furthermore, based on the results of our experiments, we could imagine biological adaptation as a *measuring worm* climbing around on the fitness landscape (Figure 14). The population of a species is represented by the worm. Its head is on the phenotypic plane and its tail is on the genotypic plane. These two planes are assumed to be correlated each other to a high degree. The landscape is always changing corresponding to the state of the worm (interactions between individuals). The worm stretches its head to the local top (first step), and when it stretches itself out, it starts pulling it's tail (second step). In our experiments, the Baldwin effect was observed once every trial. We believe that the repetition of these two steps like the behavior of measuring worms will be observed in the experiments where the environment (e.g. payoff matrix) itself is also changing. Such view of the interactions between learning and evolution might simplify the explanation of *punctuated equilibria*. In fact, Baldwin noticed that the effect might explain that variations in fossil deposits seem often to be discontinuous (Baldwin, 1896).

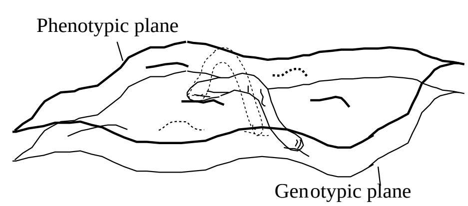

Figure 14: A "measuring worm" on the fitness landscape.

It has been also shown that implications that learning cost has on the attribution of an individual's fitness score in dynamic environments is very different from those in static environments. High evolutionary cost of learning is one of the necessary conditions for the second step of the Baldwin effect to occur in general, as pointed out by Mayley (Mayley, 1997). However, in our model the learning costs are not explicitly embed in the system. In the experiments, the second step was dominated not by time-wasting costs, energy costs, unreliability costs or so on during the vulnerable learning period. Instead, it was dominated by the constraints of the performance of the learning algorithms themselves in the complex environment where it was impossible for any algorithm to predict opponents' behavior perfectly.

The Baldwin effect generated the Meta-Pavlov [x00x] strategy, and the system stabilized with it. We have analyzed the property of the Meta-Pavlov [x00x] strategy, and have shown it's outstanding performance, which is rather a by-product to us. The excellent performance of the Meta-Pavlov [x00x] is also supported by the fact that in the extended experiments it outperformed a derivative of the Prudent-Pavlov which can be interpreted as a sophisticated offspring of the famous strategy Pavlov.

This model can be extended in several directions. One obvious direction would be to attempt to reinterpret and evaluate our results concerning the interactions between learning and evolution in the context of pure biology. Another direction would be to focus on the technical aspects of the evolutionary mechanism of varying phenotypic plasticity. It would be interesting to apply the automatic mechanism of adjusting the balance between evolution and learning in the fields of distributed AI or multi-agent systems.

References

Ackley and Littman, 1991. Interactions Between Learning and Evolution. *Proceedings of Artificial Life II*: 487–509.

Anderson R. W. 1995. Learning and Evolution: A Quantitative Genetics Approach. *Journal of Theoretical Biology* 175: 89–101.

Arita, T. 2000. *Artificial Life: A Constructive Approach to the Origin/Evolution of Life, Society, and Language* (in Japanese). Science Press.

Axelrod R. 1984. *Evolution of Cooperation*. Basic Books.

Baldwin J. M. 1896. A New Factor in Evolution. *American Naturalist* 30: 441–451.

Boerlijst M. C., Nowak M. A. and Sigmund K. 1997. The Logic of Contrition. *Journal of Theoretical Biology* 185: 281–293.

Bull L. 1999. On the Baldwin Effect. *Artificial Life* 5(3): 241–246.

Hinton G. E. and Nowlan S. J. 1987. How Learning Can Guide Evolution. *Complex Systems* 1: 495–502.

Lindgren K. 1991. Evolutionary Phenomena in Simple Dynamics. *Proceedings of Artificial Life II*: 295–311.

Mayley G. 1997. Landscapes, Learning Costs, and Genetic Assimilation. *Evolutionary Computation* 4(3): 213–234.

Maynard Smith J. 1982. *Evolution and the Theory of Games*. Cambridge University Press.

Maynard Smith J. 1996. Natural Selection: When Learning Guides Evolution. In *Adaptive Individuals in Evolving Populations: Models and Algorithms* edited by Belew R. K. and Mitchell M.: 455–457.

Menczer F. and Belew K. 1994. Evolving Sensors in Environments of Controlled Complexity. *Proceedings of Artificial Life IV*: 210–221.

Nowak M. A. and Sigmund K. 1993. A Strategy of Win-Stay, Lose-Shift that Outperforms Tit-for-Tat in the Prisoner's Dilemma Game. *Nature* 364(1): 56–58.

Sasaki T. and Tokoro M. 1999. Evolving Learnable Neutral Networks Under Changing Environments with Various Rates of Inheritance of Acquired Characters: Comparison Between Darwinian and Lamarckian Evolution. *Artificial Life* 5(3): 203–223.

Suzuki R. and Arita T. 2000. How Learning Can Affect the Course of Evolution in Dynamic Environments. *Proceedings of Fifth International Symposium on Artificial Life and Robotics* to appear.

Turney P., Whitley D. and Anderson R.W. 1996. Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect. *Evolutionary Computation* 4(3): 4–8.