論 文

Technical Papers

進化と学習の相互作用

繰り返し囚人のジレンマゲームにおける Baldwin 効果

Simulations and Analyses for an Interaction between Learning and Evolution

The Baldwin Effect in the Iterated Prisoner's Dilemma

鈴木 麗璽
Reiji Suzuki
名古屋大学 大学院人間情報学研究科
Graduate School of Human Informatics, Nagoya University, Nagoya, 464-8601, Japan.
reiji@info.human.nagoya-u.ac.jp, http://www2.create.human.nagoya-u.ac.jp/~reiji/

有田 隆也
Takaya Arita
（同 上）
ari@info.human.nagoya-u.ac.jp, http://www2.create.human.nagoya-u.ac.jp/~ari/

Keywords: baldwin effect, learning and evolution, the iterated prisoner's dilemma, artificial life.

Summary

The Baldwin effect is known as an interaction between learning and evolution, which suggests that individual lifetime learning can influence the course of evolution without the Lamarckian mechanism. Since Hinton and Nowlan showed this by computer simulation, many studies have been conducted in the static environment. Our concern is to consider the Baldwin effect in dynamic environment, especially when there is no explicit optimal solution through generations and it only depends on interactions among agents. We adopt the iterated Prisoner's Dilemma as a dynamic environment and introduce phenotypic plasticity to strategies by using a meta-learning rule termed "Meta-Pavlov". In this simulation, the Baldwin effect was observed as follows: First, strategies with enough plasticity spread, which caused a shift from defective population to cooperative. Second, these strategies were replaced by the strategy [x00x] which has a modest amount of plasticity. We have analyzed this strategy and have shown that it satisfies the ESS (Evolutionarily Stable Strategy) condition and establishes robust and cooperative relationships in the population.

1. ま え が き

進化と学習の相互作用に関する重要なトピックに Baldwin 効果 [Baldwin 96] がある．この現象は，ラマルク的な獲得形質の遺伝の仕組みが無くても，集団における個体の学習が集団全体の進化に方向性を与え，進化のスピードを促進するというものである．Hinton と Nowlan の先駆的な計算機実験によりこの効果が明確化されて以来 [Hinton 87]，多くの研究がなされてきたが，ほとんどの場合，世代を通して最適解が固定された静的な環境を前提としており，動的な環境における Baldwin 効果は未解明であった [有田 99]．

しかし，現実世界において学習が有効に働く状況を考えると，静的な環境よりもむしろ動的な環境においてその効力を発揮することから，動的な環境における Baldwin 効果の解明は重要である．

そこで本研究では，Baldwin 効果に対する一般的な解釈である学習のメリットとコストのバランス [Turney 96] に注目し，動的な環境において Baldwin 効果が確認されるかどうか，またこのバランスが進化と学習の相互作用にどのように働くかについて知見を得ることを目的とする．

動的な環境を考える場合，大きく 2 つに分けることが出来る．一つは集団が置かれた環境自体が世代を通して変化し，集団中の個体の適応度に影響を与える場合である．

Anderson は，世代を通して最適解が変動する動的な環境において学習が進化に与える影響を定量的に解析し，動的な環境においては学習にコストがかかっても学習に依存する集団へと進化することを示した [Anderson 95]．また，佐々木・所は，ニューラルネットを用いて学習する個体が，世代を通して変化する食べ物または毒を表すビット列の入力を正しく識別するようにニューロンの結合重みを学習し，適応度に応じて進化するというシミュレーションを行った [佐々木 97]．その結果，ダーウィン型の進化システムでは学習を前提として次第に動的環境自体に適応していったと報告している．両研究ともこのような動的環境における学習の重要性を示したものである．

もう一つは，動的な要因を集団自体が内包しているような場合である．たとえば，各個体の適応度が集団における個体間の相互作用に依存して決定され，世代を通し

て最適な解が決定できないような状況が考えられる．このような環境における進化と学習の相互作用に関する解析結果は，近年注目されているマルチエージェント環境による協調行動の進化的獲得などへの応用が期待できる．

我々は，後者の典型的な例として繰り返し囚人のジレンマゲームにおける戦略の進化を採用し，我々の提案するメタ・パブロフ学習に基づく可塑性を戦略の表現型に与えることで，集団全体の進化の過程に学習が与える影響を明らかにすることを目的として研究を行っている [鈴木 99a]．

本論文では，まず，進化実験においてBaldwin効果が有効に働き，協調的で安定な戦略集団へ進化したことを示し，次に進化の過程で出現し最終的に集団中を占めた安定な戦略であるメタ・パブロフ [x00x] 型戦略 [鈴木 99b] について解析する．

2. 研究の背景

2·1 Baldwin効果

Baldwin効果とは，進化と学習が相互に与える影響を，学習のメリットとコストのバランスから説明するものであり，一般には，次の2つの段階を経て，学習により獲得されていた形質が次第に生得的な形質へと進化していくものとされている [Turney 96]．

第1段階 学習により生存上有利な形質を獲得した個体が次世代に多く子孫を残す．

第2段階 十分多くの個体が生存上有利な形質を学習により獲得した集団では，学習にかかるコストのためその形質を生得的に獲得している個体が次世代に多く子孫を残す．

例えば，Ackley らは，2次元空間に作られた生態系の中で敵を回避しつつ食べ物を手に入れるタスクを，行動に対する先天的な評価基準を用いて学習する強化学習法を採用し，エージェントを進化させる実験を行った [Ackley 91]．その結果，進化の過程の初期段階においては，先天的な評価基準が優秀である個体が（学習がうまくいって）多く生き残ったが，次第に学習を必要とせず，先天的に正しい行動をとる個体群へと進化し，Baldwin効果が確認されたとしている（さらに，Baldwin効果が有効に働いた後，評価基準を構成していた遺伝子は不要になって適応度に影響を与えなくなるため，遺伝的浮動の力を大きく受けるようになったとも報告しており，これは進化と学習の相互作用における相反的な側面のひとつであると考えられ，興味深い）．

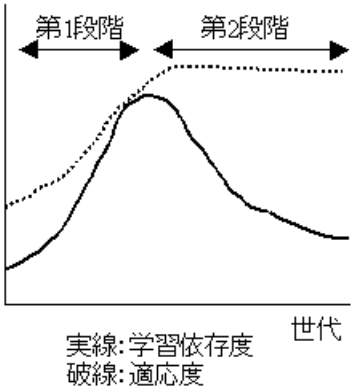


図1 適応度と学習依存度から見たBaldwin効果

このとき，集団全体の学習に対する依存度というものが定義できるとすれば，2つの段階を経て，典型的には図1のような適応度と依存度のカーブを描くと考えられる．

これまでのBaldwin効果に関する議論の多くは，効果の働く対象となる形質が世代を通して有利であることが前提とされていた．しかし，本研究では，学習によって得られるメリット（及びコスト）が世代を通して保証されないような動的な環境においても，上記のような進化と学習の相互作用が確認されれば，それをBaldwin効果として積極的に捉えることにする．

2·2 繰り返し囚人のジレンマゲーム

繰り返し囚人のジレンマゲームは，2人非ゼロ和ゲームの一種で，Axelrod による研究 [Axelrod 84] をはじめとして利己的集団における協調行動の創発に関して数多くの研究がなされている．ゲームは表1に代表される利得行列を用いて以下の手順で行われる．

- 2人のプレイヤーは協調（Cooperate）または裏切り（Defect）のどちらかの手を同時に出す．
- 出した手に応じて，利得行列（表1）から両者が得る得点が決まる．
- この対戦を繰り返し行い，その合計（平均）得点を競う．

1回きりの対戦において，プレイヤーがともに自らの期待利得を最大にするような戦略，つまりこのゲームでの支配戦略である裏切り（D）を取った場合，裏切り合いとなりこれはナッシュ均衡解である．にもかかわらず，この解はパレート最適ではなく，双方にとってより良い解すなわち協調し合いが存在するため，裏切りは正しい判断ではなかったのではないかというジレンマが生じる．さらに，十分長い繰り返しゲームにおいては，交互に裏切るよりも協調し合ったほうが双方の利益となるため，いかにして協調関係を築くことができるかが高い得点を得る際の問題となるが，協調関係を築くことができるかどうかは相手の出方次第である．つまり，ゲームにおいて

表1 囚人のジレンマゲームの利得行列

自分の手＼相手の手	協調（C）	裏切り（D）
協調（C）	(R:3, R:3)	(S:0, T:5)
裏切り（D）	(T:5, S:0)	(P:1, P:1)

（自分の得点, 相手の得点）
ただし $T > R > P > S, 2R > T + S$

ある戦略がうまくやれるかどうかは，対戦相手に大きく依存する．したがって，ジレンマゲームの戦略集団における総当たり戦の得点を適応度とするような環境を考えると，それは各個体の適応度が世代ごとに刻々と変化する動的な環境として捉えることができる．

3. 進 化 実 験

以上を踏まえ，最適解が決まらず，特に個体間の相互作用のみを考慮した動的環境として，遺伝的アルゴリズムを用いた繰り返し囚人のジレンマゲームの戦略の進化実験を取り上げ，戦略に学習（表現型の可塑性）を導入することで動的環境における進化と学習の相互作用について解析する．

3·1 戦略の遺伝子表現

各個体の持つ戦略を戦略遺伝子列 GS と学習遺伝子列 GL の 2 つの遺伝子列の組で表現する．戦略遺伝子列は Lindgren のモデル [Lindgren 91] と同様な，履歴に依存して次回の手を決定する戦略を定義する．記憶長 m の戦略は裏切りを 0，協調を 1 として以下のような 2 進数で表された履歴 h_m を持つ．

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

ここで a_0 は前回の相手の手，a_1 は前回の自分の手，a_2 は前々回の相手の手…とする．

ある履歴 k に対応して次回出すべき手を A_k（0 または 1）とすると，記憶長 m の戦略は，

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

と表すことができる．これを戦略遺伝子列とする．さらに，各 A_x に対してその表現型（協調または裏切り）が可塑性を持つかどうかを L_x（0：可塑性を持たない，1：可塑性を持つ）として，学習遺伝子列を

$$GL = [L_0 L_1 \cdots L_{n-1}] \qquad (3)$$

と定義する．例えば，しっぺ返し戦略（初回は協調，以降は前回相手が出した手を真似る）[Axelrod 84] を記憶長 2 で表すと，GS=[0101]，GL=[0000] となる．

3·2 メタ・パブロフ学習

可塑性を持つ表現型は，対戦中にその表現型を用いた結果に応じて，学習により変更される．ここで，表 2に示す学習行列を定義し，これに基づいて表現型を変更するという学習を導入する．この行列（の値）は，プレイした結果得られる得点が相対的に高ければそのまま変更せず，逆に小さければ変更するという強化学習の原理に基づくものであり，学習則としてシンプルかつ典型的なものとして今回採用する．この行列自体はパブロフ戦略（初回は協調，以降は対戦結果が相対的に良ければ次回も同じ手を出し，悪ければ手を変える）[Nowak 93] と同じ

表 2 メタ・パブロフ学習行列

自分の手＼相手の手	協調（C）	裏切り（D）
協調（C）	C	D
裏切り（D）	D	C

であるが，直前の対戦結果に応じて次回出す手を決定するのではなく，表現型を用いた結果に応じて戦略自体を変更（学習）するという意味で，この学習方式をメタ・パブロフ学習と呼ぶ．

メタ・パブロフ学習を用いた手の決定は以下の手順で行う．

- 繰り返し対戦を行う前は，各個体は GS の表す戦略をそのまま表現型として持つ．
- 表現型と履歴を参照し対戦を行い，用いた表現型（C または D）に対応する学習遺伝子列のビットが 1（可塑的）であった場合，その表現型を対戦結果に対応するメタ・パブロフ学習行列の値（C または D）と置き換えたものを新たな表現型とする．
- 次回の対戦以降，新たな表現型を参照し手を決定する．

なお，本研究では，学習すること自体にかかる明示的なコストを導入せず，学習にかかるコスト（及びメリット）はすべて個体間の相互作用の結果として与えられるものと考える．

3·3 学 習 の 例

ここで，メタ・パブロフ学習の例として，GS=[0001]，GL=[0011] の戦略が学習する例を示す．図 2（学習前）はこの戦略の表現型を図示したものである．記憶長 2 の履歴（前回の自分の手と相手の手）に対応して，次回出す手が表現型として決められている．ただし可塑性を持った表現型には下線を引いた上で，初期状態を示している．

過去の対戦履歴が CC であったと仮定すると，表現型からこの戦略は C を出す．このとき相手が D を出したと仮定する．ここで，C を出すのに用いた表現型は可塑性を持つのでメタ・パブロフ学習行列をもとに表現型を変更する．この場合，自分の手が C，相手の手が D なので，学習行列から表現型を D に変更し，次回対戦履歴が CC の場合には D を出すようになる．従って，戦略の表

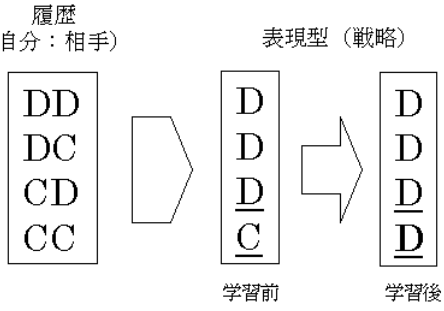


図 2 メタ・パブロフ学習の例

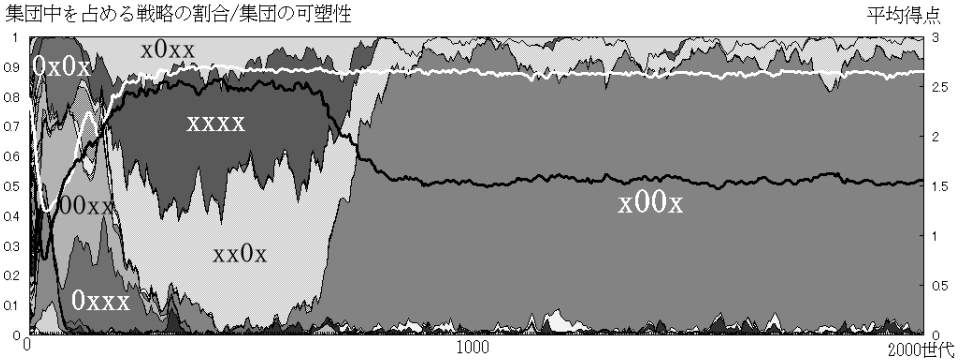


図3 実験結果（2000 世代）

現型は図2（学習後）のように変化する．

このように，学習遺伝子列に1のビットを持つ戦略個体は，繰り返し対戦を通して表現型が変化しうるという意味で可塑的な戦略であると捉える．

このように，対応する学習遺伝子のビットが1である戦略遺伝子の値は表現型の初期値としてのみ働く．そこで，今後各戦略を，可塑性を持つ学習遺伝子に対応する戦略遺伝子を x と置き換えた戦略遺伝子列でまとめて表現することで，進化の過程の大枠を捉えることにする（e.g. GS = [1000]，GL = [1001] → [x00x]）．

3·4 繰り返し対戦

以上のような戦略個体同士でノイズありの繰り返し対戦を表1の利得行列を用いて行う．ノイズとは，繰り返し対戦において，各戦略個体が出すべき手が一定の確率で反転してしまうことで，現実世界における表現の間違い，転送経路のノイズ，誤解などの不可抗力を象徴するものである．

3.1 節で示したとおり，本研究で用いられる戦略が手を決定するためには履歴が必要である．そこで，各繰り返し対戦の一番初めは，繰り返し対戦ごとにランダムに作成された仮想の履歴を各個体が参照し，初回の手を決定するものとする．

繰り返しゲームを行う状況として,「十分長い間繰り返されるが，実際何回繰り返して行われるかはプレイヤーには分からない.」という設定にするため，繰り返しの回数は固定せず，対戦ごとに一定の確率で次回の対戦が行われるものとする．この確率を未来係数と呼ぶ.

また，可塑的な戦略における表現型は，各繰り返し対戦ごとに初期状態（戦略遺伝子列が示すままの状態）に戻されるものとする．

3·5 遺伝的オペレーション

上記のような繰り返し対戦を集団全体において総当たりで行い，その合計得点を各戦略個体の適応度とする．つづいて，各適応度に応じたルーレット選択により次世代の集団を生成する．その際，一定の確率で遺伝子のビットが反転する，一点突然変異を導入する．

なお，計算量を軽減するために，はじめて行う対戦カードの場合は，繰り返し対戦を20回行った平均得点を用いるとともに保存し，すでに行ったことのある対戦カードでは保存した得点を利用するものとする．また，保存した得点は500世代ごと消去し，新たに計算し直すものとする．

4. 実験結果と考察

4·1 実験結果

記憶長2（初期集団における各遺伝子の値はランダム）の集団において，パラメータとして突然変異率1/1500，個体数1000，ノイズは起きるがそれほど頻繁ではない設定としてノイズ率1/25，繰り返し対戦が十分長い状況として未来係数99/100，世代数2000を用いて進化実験を行った．実験結果の一例として図3，図4を示す.

ここで集団の可塑性（黒実線）とは，学習遺伝子列中に占める1のビットの割合を示し，これは2·1節における学習依存度に相当する指標と捉えることができる．また平均得点（白実線）は各世代に行われたすべての対戦の得点を平均したもので，互いに協調し合ったときに最も高くなる（3点）ことから協調の度合いを表す指標として捉える．

この試行における進化の過程の概略を示す．はじめの約60世代までは，裏切り的な戦略（[0000]，[000x] など）が平均得点を低下させた．またそれとほぼ同時に [0x0x]，[00xx]，[0xxx] などの一部可塑的な戦略も集団中を占めた．その後，集団の可塑性の増加とともに [xxxx]，[x0xx] といった可塑的な戦略が集団中を占め協調関係を築き，約200世代までに高い平均得点を持つ集団へと進化した．これまでの過程で，可塑性は裏切り的な集団から協調的な集団へのシフトに有利な方向へと働いたと考えられ，これはBaldwin効果の第1段階と捉えられる．

その後，高い平均得点を維持したまま，集団の可塑性は次第に低下し約50%のところで安定し，最終的には集団の大部分を [x00x] 型の個体が占める結果となった．これは，一部の可塑性がコストとして働いて，集団を維持

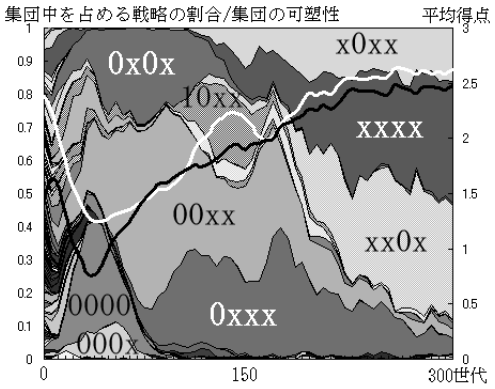


図 4　実験結果（300 世代）

するのに最低限必要な可塑性を持った戦略が選択されたためと考えられ，Baldwin 効果の第 2 段階と捉えられる．

行った試行のほとんどで，最終的に [x00x] 型の戦略が集団中を占めた．また，この結果に見られるような傾向をもつ進化の過程が，試行の約 70% で確認された．

また，試行の残り 30% の中には，第 1 段階の後，可塑的な集団から第 1 段階の初期の裏切り的な戦略が集団中を占める状態に戻ることが何度か繰り返される場合があった．過度に可塑的であることが裏切り的な戦略の侵入を許した結果，[x00x] への進化を遅らせたことがこの原因と考えると，進化と学習の相反的側面と捉えることができ興味深い．このほか，初期段階から直ちに [xxxx] などの可塑的で協調的な戦略が集団中を占めた場合，Baldwin 効果が観察されず直接 [x00x] 型戦略が集団中を占める場合などが観察された．

なお，ノイズ率を変えて実験を行うと，ノイズ率が高いほど裏切り的な戦略（特に [0001] 型戦略）が集団中に広まりやすい傾向が見られた．

4·2　戦略の推移と Baldwin 効果

4·1 節の結果から，このような動的な環境においても Baldwin 効果と捉えられる進化の過程が確認されたが，このような過程において，可塑性は戦略の特徴にどのような影響を与えたのか，戦略を特徴ごとに分類して解析する．

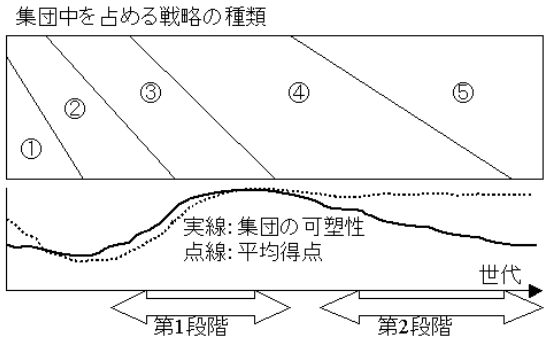
進化の過程で集団中を占めた戦略を，その特徴ごとに分類すると図 5のようになる．ただし，実際にはこれほどはっきりとした遷移にはなっておらず，グループごとの境界はあいまいである．また，遷移のスピードも試行によって異なる．

はじめに，初期集団においては裏切り的な戦略に対して搾取される戦略①が多く含まれるため，②のような全面裏切り的な戦略が有利となり，集団中に広まる．②を中心とした戦略が集団中を占めると，対戦が裏切り合いばかりになり平均得点は低下する．この状態では，③のような戦略が徐々に集団中に広まる．これらの戦略は②のような戦略と同様に一度裏切り合ったら裏切り合いを続けるが，同種同士の対戦では戦略に含まれる可塑性によってノイズや初期状態をきっかけにして協調を出し合うので，裏切り合いと比べて若干高い得点を得るためである．

その後，③が増加し，集団全体においてさらに協調する機会が増え，平均得点が上昇してくると，裏切り的な戦略に対してそれほど点を与えないことを維持しつつも，③と比べてより協調的であることで高い得点を得る④のような，より可塑的な戦略が集団中を占めるようになる．

ここで，最終的に集団の大半を占める⑤よりも④の方へと進化する傾向が高いのは興味深い．③から④への進化のように平均得点が上昇している状況では，他種の得点を下げるよりもまず自身の得る得点が高いことが集団中に広まるために必要とされる．しかし，⑤の ESS 的な性質（5·1 節で後述）が③との対戦で両者の得る得点を，④と比べて下げてしまっているので，④の戦略が先に集団中を占めると考えられる．このことは，ESS 的な戦略であるからといって，他の戦略集団に容易に入り込むことができる訳ではないことを示している．これまでの可塑的な協調集団に至る過程が Baldwin 効果の第 1 段階である．

その後，④のような協調的な戦略ばかりになると，対戦のほとんどが協調し合いになり，異なる戦略同士の適応度の差が小さくなる．この状態においてほとんどの対戦は，基本的に協調しあい，ノイズが入ると一定の回復過程を経て協調しあいにもどるというサイクルになるため，戦略の差はノイズが入ってからの振る舞いに現れる．このとき，ノイズをきっかけに自分にとって不利な協調を出す可能性のある余分な可塑性，すなわちコストとして働く可塑性が取り除かれていき，最終的には「協調の維持に必要最小限の可塑性を持った戦略」⑤が徐々に集団中を占める．[x00x] は記憶長 2 の戦略群においては ESS 条件を満たすため（5·1 節で後述），ここで集団は安定す



①: 全面協調的 ([1111], [1101]など)
②: 全面裏切り的 ([0000], [000x]など)
③: 裏切り中心だが一部可塑的で協調的
([00xx], [0x0x], [0xxx]など)
④: 可塑的で協調的 ([xxxx], [x0xx], [xx0x]など)
⑤: 集団の維持に適度に可塑的で協調的 ([x00x])

図 5　進化の過程における戦略の分類

る．この過剰な可塑性の減少と安定化が，Baldwin 効果の第 2 段階である．

以上から，この過程において，学習は裏切り的な戦略集団から協調的な戦略集団へのシフトと安定化に大きな影響を与えていると考えられる．また，戦略における可塑性は，対戦する戦略に応じて自身の振る舞いをうまく変える特徴を実現するのに有効であると考えられる．

5. メタ・パブロフ [x00x] 型戦略の解析

実験結果から，最終的に集団中を占めたメタ・パブロフ [x00x] 型戦略は非常に安定であることがわかった．そこでこの戦略について，戦略の安定性や特徴，学習の果たす役割について解析する．

5·1 ESS 条 件

集団における戦略が進化の過程で安定であるかどうかの基準として，Maynard-Smith が提案した「進化的に安定な戦略（ESS）[Maynard-Smith 82]」がある．集団において ESS を満たす戦略 a の条件は，

$E(a,b)$ を戦略 a と b との対戦で a が得る得点とすると，

$$E(a,a) > E(b,a) \qquad (4)$$

または

$$E(a,a) = E(b,a) \text{ かつ } E(a,b) > E(b,b) \qquad (5)$$

が他のすべての種類の戦略 b に対して成り立つことである．

[x00x] がこの条件を満たすかどうか確認するために，[x00x]（GS=[0000]GL=[1001]）と記憶長 2 の可能なすべての戦略 256 個との繰り返し対戦をノイズ率 1/25，未来係数 99/100 で 100 回行ったときの各対戦の平均得点を計算した．

図 6は [x00x] と記憶長 2 の戦略との対戦成績である．横軸は，各戦略の遺伝子列を [GSGL] とならべて 8 ビットの 2 進数として見た場合の値を表す．縦軸は，[x00x] 同士の対戦で [x00x] が得た得点と各戦略個体と [x00x] との対戦において各戦略個体の得た得点の差，すなわち

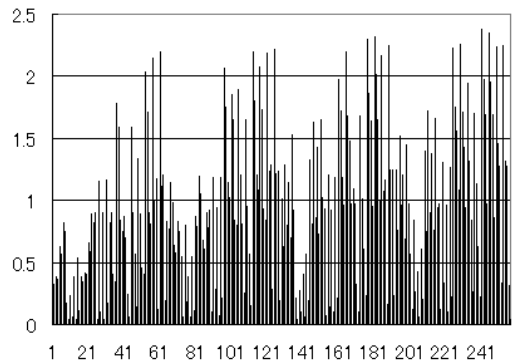


図 6 [x00x] 型戦略と記憶長 2 の戦略との対戦成績

(4) 式の左辺と右辺の差である．したがって，[x00x] が ESS であるには縦軸の値がすべて 0 より大であれば良いが，図 6からこの試行においては ESS 条件を満たしていると言える．

5·2 状態遷移モデル

メタ・パブロフ [x00x] 型戦略は，2 つの可塑性を持った表現型を内部状態とすることで，状態遷移モデルとして表現できる．

図 7に [x00x] 型戦略の状態遷移を示す．各矩形は自分と相手の出した手，および可塑性を持った表現型の状態を示す．上下に並んだ 0 または 1 は，上が相手の出した手，下が [x00x] の出した手を表す（0=裏切り，1 =協調）．[x00x] の出した手につけられた添え字は，可塑性を持った表現型の現在の状態を示し，順に [x00x] の前の x と後ろの x の表現型の状態を示す．各状態から 2 本ずつ伸びた矢印は，相手の取りうる手に依存して可能な状態遷移先を示している．

任意の戦略と [x00x] との繰り返し対戦は，相手の戦略に応じて矢印を選んでいくことで表現できる．たとえば，[x00x] と全面裏切り戦略 [0000] が対戦した場合，相手の手が常に 0 となるように矢印を選んでいけばよい．この場合，たとえばサイクル②をとり続けることになる．

この図において注目すべき点は，協調関係を持続する状態（状態 A）が崩れたとき，もう一度もとの状態 A に最短で戻るには，相手が「裏切り，協調，裏切り」という複雑な手（サイクル①）を取らなければならないことである．実は,「裏切り，協調，裏切り」のあと協調に戻るという協調の回復過程は，[x00x] 同士の対戦において実現される．つまり，協調関係が崩れたときに最も早く協調関係を回復することのできる相手の一つが同種であるということである．逆に言えば，同種以外の戦略に対して，[x00x] は協調関係を回復しにくいことを表してお

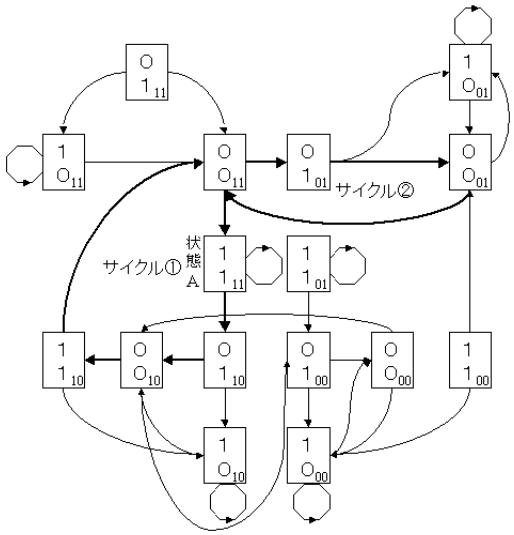


図 7 [x00x] 型戦略の状態遷移

表 3 全面裏切り戦略と [x00x] との対戦例

0000	000000000000 …平均約 2.3
x00x	010010010010 …平均約 0.8

表 5 [x00x] 同士の対戦での協調の回復例（下線はノイズ）

x00x	1111<u>0</u>0101111 …
x00x	111110101111 …

り，この特徴は [x00x] の ESS 的な性質を生み出すメカニズムの一つであるとみなすことができる．

5·3 学習の役割

これまで，進化の過程において，過剰な可塑性が減少した結果，最終的に [x00x] 型の戦略が集団中を占めたと論じてきたが，ではどうして,「前の x」と「後ろの x」という 2 つの可塑性を残して減少は止まったのだろうか．状態遷移モデルからわかるとおり，戦略の特徴は遺伝子列全体を総合して決まるが，ここではあえて前の x，後ろの x，その両方それぞれの意義について検討する．

前の x は裏切り中心の戦略に対して，多く点を取らせず，侵入させないという点で有効であると考えられる．図 7において，サイクル②は全面裏切り戦略と対戦した場合にたどるものの一つである．このとき，前の x の可塑性により [x00x] は 2 回に 1 回裏切られる．[x00x] の得る得点は約 0.8 点と低いものの，このおかげで全面裏切り戦略が得る得点は約 2.3 点に下がる（表 3）．[x00x] 同士の対戦で [x00x] が得る得点は約 2.6 点であるから，これは [x00x] を進化的に安定な戦略にするのに貢献している．

一方，後ろの x は協調中心の戦略に対して，ノイズによる偶発的な裏切りをきっかけにして裏切りに転じる点で有効であると考えられる．図 7中央の互いに協調し合った状態から，協調が崩れるとき，後ろの x は 1（協調）から 0（裏切り）に変わる．これは次回協調し合った後，裏切ることを示しており，これは協調を維持するような戦略に対して搾取する (表 4) とともに，相手に点を取らせない点で有効である．

以上のとおり，前後の x にはそれぞれ裏切り的・協調的戦略に点を与えない働きがあると考えられる．しかし，この働きを生かすには，[x00x] 同士の対戦で必ず強固な協調関係が築かれなければならない．前述の通り，[x00x] はサイクル①によってノイズに対してうまく協調を回復するが，この遷移の中では前後の x についてそれぞれ 2 回ずつ交互に学習（結果的に状態が変化しない学習も含む）が行われている（表 5）．これは前後の x による学習が共同してうまく働いた結果，強力な協調関係を築いていることを示している．

6. むすび

本研究では，動的環境における進化と学習の相互作用を解析するため，個体間の相互作用のみに依存した動的環境である繰り返し囚人のジレンマゲームの戦略の進化に表現型の可塑性を導入して，実験を行った．

その結果，このような動的な環境においても Baldwin 効果が観察され，集団は適度な可塑性を持った安定な協調集団へと進化した．このとき，集団は直接協調的で安定した集団へと進化するのではなく，一旦集団の可塑性が増加する方向へ進化し十分な協調関係を実現してから，可塑性の減少とともに協調的で安定な集団へと進化した．

さらに，最終的に集団の大部分を占めたメタ・パブロフ型 [x00x] 戦略の解析を行ったところ，[x00x] 型戦略は必要最低限の可塑性を持った強力な戦略であることが判明した．これは [x00x] 型の戦略が純粋にジレンマゲームにおける強力な戦略として興味深いだけでなく，このような進化と学習の割合が自動的に調節される進化の枠組みが，動的な環境に対する集団の安定性を実現するための重要なファクターとなっていることを示唆する点で意義深いと考えられる．現在，[x00x] 型戦略について解析を続けるとともに，記憶長を突然変異によって変化させたり，学習方式そのものを遺伝子に表現して進化させたりするようなオープンエンドな進化実験を行っている．また，可塑的な形質の割合が進化的に調整されるこのような仕組みの工学的応用についても検討している．

表 4 [x001] と [x00x] の対戦例（下線はノイズ）

x001	1111<u>0</u>01101101 …
x00x	1111101001001 …

◇ 参 考 文 献 ◇

[Ackley 91] Ackley, D. , Littman, M. : Interaction Between Learning and Evolution, Artificial Life II, pp. 487–509 Addison-Wesley (1991).

[Anderson 95] Anderson, R. W. : Learning and Evolution: A Quantitative Genetics Approach, Journal of Theoretical Biology, Vol. 175, pp. 89–101 (1995).

[有田 99] 有田隆也: 人工生命, 科学技術出版 (1999).

[Axelrod 84] Axelrod, R. : The Evolution of Cooperation, Basic Books, NewYork (1984).

[Baldwin 96] Baldwin, J. M. : A New Factor in Evolution, American Naturalist, Vol. 30, pp.441–451 (1896).

[Hinton 87] Hinton, G. E. and Nowlan, S. J. : How Learning Can Guide Evolution, Complex Systems, Vol. 1, pp. 495–502 (1987).

[Lindgren 91] Lindgren, K. : Evolutionary Phenomena in Simple Dynamics, Artificial Life II, pp. 295–311 Addison-Wesley (1991).

[Maynard-Smith 82] Maynard-Smith, J. : Evolution and the Theory of Games, Cambridge University Press (1982).

[Nowak 93] Nowak, M. A. and Sigmund, K. : A Strategy of Win-Stay, Lose-Shift that Outperforms Tit-for-Tat in the Prisoner's Dilemma Game, Nature, Vol. 364, No. 1, pp. 56–58 (1993).

[佐々木 97] 佐々木貴宏, 所 真理雄: 進化的エージェント集団の動的環境への適応, コンピュータソフトウエア, Vol. 14, No. 4, pp. 33–46 (1997).

[鈴木 99a] 鈴木麗璽, 有田隆也: 囚人のジレンマゲームにおける Baldwin 効果, 人工知能学会第 13 回全国大会論文集，pp.277–278 (1999).

[鈴木 99b] 鈴木麗璽, 有田隆也: メタ・パブロフ：進化と学習による適応を自動調節する繰り返し型囚人のジレンマ戦略, 情報処理学会研究報告, 99-GI-1, pp.15–22 (1999).
[Turney 96] Turney, P. , Whitley, D. and Anderson, R. W. : Evolution, Learning, and Instinct: 100 Years of the Baldwin Effect, Evolutionary Computation, Vol. 4, No.3, pp. 4–8 (1996).

〔担当委員：××○○〕

19YY 年 MM月 DD日　受理

著　者　紹　介

鈴木　麗璽（学生会員）

1998 年名古屋大学情報文化学部自然情報学科退学（飛び級）．現在，名古屋大学大学院人間情報学研究科博士課程（前期）在学中．

有田　隆也（正会員）

1983 年東京大学工学部計数工学科卒業．1988 年同大学大学院工学系研究科修了．工学博士．名古屋工業大学講師，カリフォルニア大学ロサンゼルス校客員研究員を経て，現在，名古屋大学大学院人間情報学研究科助教授．人工生命や情報科学の研究に従事．言語の進化，人間行動の進化，進化的計算論などに興味を持つ．情報処理学会，日本認知科学会，電子情報通信学会各会員．