修士論文

動的環境における
進化と学習の相互作用に関する研究

学籍番号：３１９８０１３５７

鈴木麗璽

名古屋大学　大学院人間情報学研究科

物質・生命情報学専攻

1999 年度

目次

1.はじめに

進化と学習の相互作用に関する重要なトピックに，Baldwin 効果と呼ばれる現象がある．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，その結果学習によって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というものである．

これまで，Baldwin 効果に関する進化実験による研究は，その存在を明確にした Hinton と Nowlan による先駆的な進化実験をはじめとして最適解が固定されたものがほとんどであり，動的な環境において Baldwin 効果がどのように働くかは未解明であった．しかし，現実世界においては，学習はむしろ動的な環境において有効に働くと考えられるため，動的な環境における進化と学習の相互作用を明らかにすることは重要であると言える．

そこで本研究では，特に個体間の相互作用にのみ適応度が依存した動的な環境として繰り返し囚人のジレンマゲームの戦略の進化を取り上げ，戦略に提案するメタ・パブロフ学習に基づく表現型の可塑性を導入した進化実験を行い進化の過程を解析することで，動的な環境における進化と学習の相互作用について知見を得ることを目的とする．

本論文では，まず基本的な進化実験において，このような動的な環境においても 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す．次に，戦略の推移と Baldwin 効果の関係，および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]型戦略について解析する．さらに，モデルをより一般化したオープンエンドな進化実験の結果について示し，考察する．

本論文の構成は次のとおりである．２章では，研究の目的と Baldwin 効果について述べる．3 章では Hinton と Nowlan による Baldwin 効果に関する先駆的な進化実験および進化と学習の相互作用に関する関連研究を紹介する． 4 章では構築したモデルについて詳細に解説する．5 章では基本的な進化実験の結果および戦略の推移の解析について示し，メタ・パブロフ[x00x]型戦略についての解析を 6 章で行う．7 章ではモデルをより一般化した進化実験の結果について示す．最後に 8 章では，これまでの結果および考察をまとめるとともに，今後の研究の発展の可能性について述べる．

2.研究の目的

2.1 Baldwin 効果

進化と学習の相互作用について，様々な議論がなされてきた．そのうち最も大きなトピックのひとつに獲得形質の遺伝の是非をめぐる議論があった．その中で，Lamarck 的な獲得形質の遺伝の仕組みが無くても，自然選択のみでそれと同様の効果が得られるというシナリオが Baldwin によって示された．

Baldwin 効果は，およそ 100 年ほど前，Lamarck 的な獲得形質の遺伝の仕組みを用いず，自然選択のみによって進化と学習の相互作用を示すものとして Baldwin によって提案された（Baldwin 自身はこれを Organic Selection と呼んだが，その後 Baldwin 効果という名で定着した）[Baldwin 1896]．これは，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，現在の一般的な定義では，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, Whitley and Anderson 96].

なお，ここで述べる学習とは，人間の知能に代表されるような高度な学習メカニズムだけでなく，たとえば運動による筋肉の増強や日焼けによる皮膚の色の変化などを含む表現型の生涯の変化，すなわち表現型の可塑性を示すものである．

- 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．

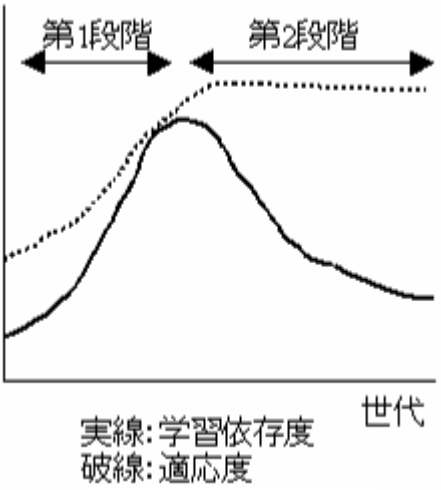

図 1：Baldwin 効果

第 1 段階は，学習によるメリットが中心的な選択圧として働いた状態，第 2 段階はコストが選択圧として働いた状態である．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．

Baldwin によってこの効果が提案されたのは非常に昔のことであるが, その存在を確認するのが難しいために, 長い間あまり注目されていなかった. しかし近年, Hinton と Nowlan による先駆的な進化実験（3.1 節参照）によりこの効果が明確にされて以来，生物学的側面からだけでなく，進化的計算などの工学的分野からも注目されるようになり，新たな局面を迎えている．

2.2 動的な環境における進化と学習の相互作用

ところで，これまで Baldwin 効果に関して行われた進化実験は，Hinton と Nowlan による先駆的な進化実験[Hinton and Nowlan 87]をはじめとして最適解が固定されたものがほとんどであり，これまで動的な環境において Baldwin 効果がどのように働くかは未解明であった.

しかし，3.1 節で示したような進化と学習の相互作用が起きるために，環境が静的であることは，必ずしも必要ではない．なぜなら，Baldwin 効果の働く要因である学習によるメリットとコストは, 動的環境においても存在するからである. たとえば, 刻々と変化する状況に対し，学習は柔軟に振る舞う術を与えてくれる．また，学習は常に良い結果ばかりをもたらすわけではなく，環境の変動が誤った学習や不必要な学習を引き起こす状況も考えられる．したがって，動的な環境においても，学習のメリットやコストが働き Baldwin 効果のような現象が起きることは十分期待できる．また，現実世界において，学習は不安定な環境に柔軟に適応するために不可欠であることを考えると，動的な環境での議論はより現実的な設定であると言える．

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[鈴木 99], [鈴木，有田 99a，99b，99c].

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, もう一つは動的な要因を集団の各エージェント自体が内包しているような場合である．これまで，前者に注目した進化と学習の相互作用についての研究はいくつかあったが（3.2 節参照），後者についての Baldwin 効果に関する研究はほとんどなされていない．本研究では後者のような環境を研究の対象とする．このような環境として，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．

本研究ではその典型的な例として，戦略に表現型の可塑性を導入した繰り返し囚人のジレンマゲームの戦略の進化モデルを構築し，進化実験とその解析を行う[Suzuki and Arita 2000]，[鈴木，有田 2000]．その際，次のような点に注目する.

1. **戦略に学習を導入することで, このような動的な環境においても** Baldwin **効果は確認できるだろうか.**
2. **囚人のジレンマゲームに関する議論の焦点のひとつである，戦略集団における協調関係の創発に，学習はどのような影響を与えるか.**

3．その結果あらわれるジレンマゲームの戦略とはどのようなものか.

3. 関連研究

3.1 Baldwin 効果に関する先駆的研究

Hinton と Nowlan は Baldwin 効果を確認するために，遺伝的アルゴリズムを用いたシンプルなモデルを用いて進化実験を行い，特に Baldwin 効果の第 1 段階について明快に示した[Hinton and Nowlan 87]. 彼らが行った進化実験とは次のようなものである．

集団における各個体は 1 または 0，？からなる 20 個の遺伝子の列を持つ．初期集団（1000 個体）における各個体の遺伝子は 1:0:?=1:1:2 の比率でランダムに決定される．各個体はネットワークを表し，各遺伝子はネットワーク内の結合のパターンを表す．結合には正しい結合（1）と誤りの結合（0）があり，遺伝子が 0 または 1 であることは，結合がそれぞれ遺伝的に 0 または 1 に決定されていることを示す．遺伝子が？であることは結合が遺伝的には未決定であることを示し，この部分における正しい結合を学習作業によって後天的に獲得する． 学習作業および適応度の計算方法は次のとおりである．

A) 遺伝子中の"?"すべてについて，1 または 0 をランダムに当てはめる．

B) 正しいネットワークの結合（この場合はすべて 1）と比較し，すべての結合が正しい結合と一致すれば，学習が完了したものとして C へ移る．そうでなければ A に戻る．このとき i を学習が完了するまでに行った試行数とする．ただし最大試行回数の 1000 回まで試行しても解と一致しない場合は i=1000，初めから解と一致している場合は i=0 とする．

C) 以下の式を用いて適応度を計算する．

$$\text{適応度} = 1 + \frac{19(1000 - i)}{1000} \qquad (1)$$

（1）式から，すべての結合が遺伝的に正しい個体（の適応度は最大値 20，遺伝的に誤りを含む個体の適応度は最小値 1，正しい結合を学習によって獲得する可能性のある個体の適応度はこの間の値をとることになる．

適応度に応じたルーレット選択で親 2 個体を選び，一点交叉してできた子孫のうち片方を次の世代の個体とする．これを 1000 回繰り返し，次世代の集団を生成する．

以上のようなモデルで実験を行ったところ，図 2 ような結果が得られたと報告している．彼らは，全遺伝子中に占める 0，1，？の割合の推移に注目した．初期数世代まではすべての個体が学習に失敗し，選択圧がかからない状態が続いた．その後，学習によって正しい結合の獲得に成功した個体が出現し，集団中に広がり始めた．

このとき, 全遺伝子中に含まれる 0 の割合は急激に減少し, 1 と？の割合は上昇した. これが学習によるメリットが働いた状態で, Baldwin 効果の第 1 段階である. その後, ？の割合が若干減少したところで, 進化は収束した. この？の若干の減少と 1 の割合の上昇が Baldwin 効果の第 2 段階である.

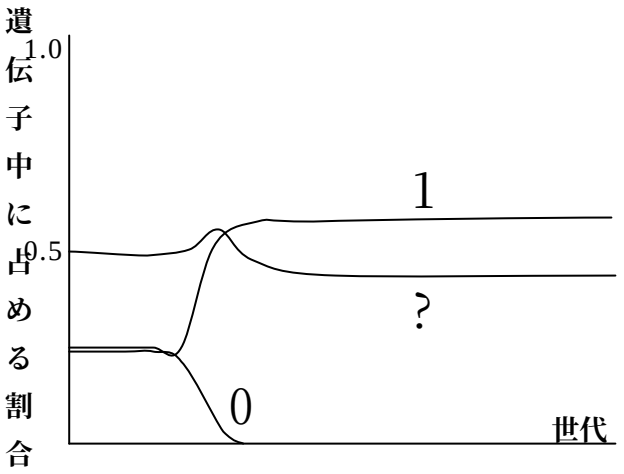

図 2：Hinton と Nowlan の実験結果（概要）

この実験において注目すべき点は, 学習作業の結果が次世代に遺伝する仕組みが無いにもかかわらず, 学習が集団の進化に大きな影響を与えている点である. この結果では最終的に遺伝的に正しい結合の割合が約 60%もの集団へと進化したが, 仮に, この実験において学習を導入しなかった場合, すなわち遺伝子の値として 0 または 1 だけを用いる設定で実験を行った場合, 唯一高い適応度を持ったすべての遺伝子が 1 である個体を $(1/2)^{20}$ の確率で発見しなくてはならないため, 選択圧がかからないままの状態が続くことは容易に想像できる. この結果から彼らは, 学習結果が次世代に伝わらなくても最適解への進化を促進する学習のこのような働きを, 「尖った適応度地形をなだらかにする」役割があると指摘している.

なお, この実験では Baldwin 効果の第 2 段階が大きく現れていない. この理由は, このモデルでは突然変異を用いておらず, 第 2 段階が十分働く前に集団が 1 種類の遺伝子に収束し, 進化が止まったためであると推測される. 実際, Harvey や我々の行った突然変異を導入した追実験では, 第 2 段階は継続され, 徐々に正しい結合を多く持った集団へと進化していった[Harvey 96], [有田 2000].

この実験結果は, Baldwin 効果の存在をはじめて明確に示したのと同時に, 進化と学習が組み合わさったメカニズムの有用性を示したものとして重要である.

3.2 進化と学習の相互作用に関する研究

Ackley らは，2 次元空間に作られた生態系の中で，エージェントが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（Evolutionary Reinforcement Learning）を採用し，集団を進化させる実験を行った[Ackley and Littman 91]．各エージェントは外部からの視覚情報に対してとるべき行動を出力するニューラルネットワークと，入力情報に対して正または負の評価を与えるニューラルネットワークを持ち，その結合重みの初期値が遺伝子として与えられる．各エージェントは行動を出力するネットワークに従って行動するが，その際評価基準を与えるネットワークの評価を強化信号として，バックプロパゲーション法により行動を出力するネットワークの結合重みを更新する．実験の結果，進化の過程の初期段階においては，正しい評価を与えるネットワークを持つ個体が，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報告している．また，彼らはさらに，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も考えられ，興味深い．

この研究は，学習に関してより具体的なモデルを用いて，Baldwin 効果を示した点で重要である． また，強化学習の分野において，学習の効率に大きな影響を与える強化信号をいかにして与えるかが問題となっているが，ERL は強化信号を与える規準自体もエージェントに内包させ，進化の枠組みに取り込んでいる点で興味深い．しかし，世代を通して最適解が固定されている（敵を避けて食べ物を手に入れるエージェントが最適）という意味では，本研究で採用するモデルとは異なるものである．

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遺伝的な流れに与える影響を，力学系の方程式を用いた解析的な手法で定量的に分析した．その結果，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]．彼はまた，個体の学習には遺伝的な多様性を維持する働きがあり，これが環境の変動に追従することを可能にしているとも主張している．

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変更し，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，所 97]．このとき，学習結果（すなわち学習後の結合重み）が次世代の個体に遺伝するラマルク型の進化システムと，学習結果が遺伝せず学習前の初期値が

遺伝するダーウィン型の進化システムを用いて実験を行ったところ，ラマルク型の進化システムでは動的環境に追従しきれなかったが，ダーウィン型の進化システムでは進化の過程で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．

両研究とも動的環境における学習の重要性を示したものであるが，動的な要因が環境自体にあるという点では，本研究の対象とは異なるものである．

4.モデル

以上を踏まえ，世代を通して最適解が決まらず，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モデルを構築した．繰り返し囚人のジレンマゲームついて簡単に説明した後，モデルについて解説する．

4.1 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．

表 1：囚人のジレンマゲームの利得行列

相手の手（→） 自分の手（↓）	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)

2R>T+S

（自分の得点，相手の得点）

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（2R>T+S）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．

4.2 戦略の遺伝子表現

集団における各エージェントは繰り返し囚人のジレンマゲームの戦略を遺伝子として持つ．このモデルでは，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．

$$h_m = (a_{m-1}, \cdots, a_1, a_0)_2 \qquad (2)$$

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，

$$GS = [A_0 A_1 \cdots A_{n-1}] \quad (n = 2^m) \qquad (3)$$

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

$$GL = [L_0 L_1 \cdots L_{n-1}] \qquad (4)$$

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．

4.3 学習規則

可塑性を持つ表現型は，繰り返し対戦中に学習則によって変更される．本研究では 2 つの学習則を用いて表現型を変更する．一つはランダム型学習，もう一つは学習行列を用いたメタ・パブロフ学習である．学習は次の手順で行われる．

- 繰り返し対戦を行う前は，各個体は GS の表す純粋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，以下のどちらかの手順で表現型を書き換える．

  １．ランダム型：その表現型をランダムに C または D と置き換える．

  ２．学習行列型（メタ・パブロフ学習）：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．

ランダム型学習は，対戦の結果によらずに表現型を変更する点で学習とは言い難いが，この研究では表現型の可塑性を学習の根本として考えていることから学習の一つとして捉える．また，学習行列型では標準的な学習行列の値として表 2 に示すメタ・パブロフ学習行列を定義し，これに基づいて表現型を変更する．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 and Sigmund 93]と同じであるが，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，この行列を用いた学習方式をメタ・パブロフ学習と呼ぶ．

表 2：メタ・パブロフ学習行列

相手の手（→） 自分の手（↓）	C	D
C	C	D
D	D	C

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 3（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，戦略の表現型は図 3（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．

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]）．

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．

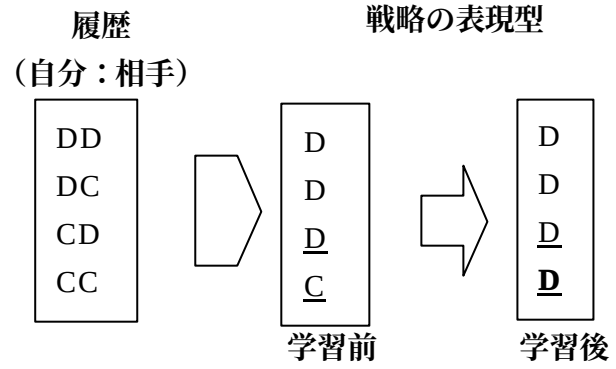

図 3：メタ・パブロフ学習の例

4.4 繰り返し対戦と進化

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．

4.2 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，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.

こ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．最後に，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.

5. 基本的な進化実験

以上のモデルを用いて行った進化実験の結果について示す．はじめに，基本的な実験として記憶長 2 の戦略群を用いて学習無しの場合，学習を導入した場合について実験を行った．

5.1 学習なしの戦略での実験

戦略に学習を導入しない場合の集団の挙動を観察するために，各戦略の学習遺伝子 GL を常に[0000]に固定した上で，GS の各ビットをランダムに選んだ記憶長 2 の戦略群を初期状態として，進化実験を行った．

なお，パラメータとして突然変異率 *1/1500*，個体数 *1000*，ノイズ率 *1/25*，未来係数 *99/100*，世代数 *2000* を用いた．今後，特に断りの無い限り，各実験においてこれらのパラメータを用いる．

進化実験の典型的な結果を図 4 に示す．世代を通して塗りつぶされた領域は集団中を占めた遺伝子の分布を示す．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．

[0101]（しっぺ返し戦略）→[1001]（パブロフ戦略）→[0001]→[0101]→…というサイクルに代表されるような，裏切り的な戦略と協調的な戦略が交互に集団中を占める状態が続き，平均得点は大きく振動した．この結果は Lindgren の報告した進化の過程と類似したものである．

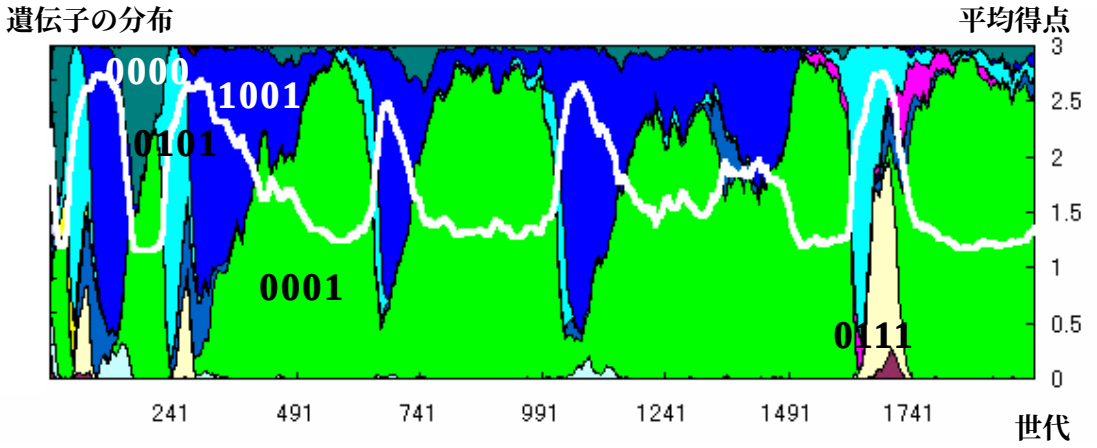

図 4：記憶長 2 学習なしの実験結果

5.2　記憶長 2 ランダム型学習

次に，記憶長 2（初期集団における GS，GL 各遺伝子の値はランダム）の集団において学習方式をランダム型学習として同様のパラメータを用いて進化実験を行った．

実験結果の一例を図 5 に示す．ただし，試行ごとに進化の挙動は大きく異なることを断っておく．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して捉えることができる．

この試行では，初期集団から裏切り的な戦略が集団中を占め平均得点が低下した．つづいて若干の集団の可塑性の増加を伴いながら [x0x1]や[01x1]という戦略が集団中の大半を占め平均得点の高い集団へと進化した．これらの戦略は協調し合いを維持するが，ノイズなどにより協調関係が崩れると，その可塑性によって確率的に協調関係を回復する戦略である．その後，可塑性は低下し，[x001]型の戦略が長い世代にわたって集団中を占める状態が続いた．この戦略は，同種同士の対戦において基本的に協調し合い，ノイズが入って協調関係が崩れると，可塑性により確率的にではあるがより早い段階で協調し合いに戻ることの出来る戦略である．しかし，集団はこのまま収束する訳ではなく，さらに可塑性は低下し，より協調的な[1001]型の戦略が徐々に侵入してきた．この[1001]型戦略が増加する方向への進化は集団全体の性格をより協調的に変化させるため，裏切り的な [0001]型戦略の侵入を許し，平均得点の低下を招いた．

以上のうちで[x001]→[1001]→[0001]型戦略のような過程は他の試行においても多く確認された．[x001]型戦略は他の試行においても集団中を長い期間占めたことから，比較的安定な戦略であると言える．しかし，集団が完全に収束するに至らなかったのは，記憶長 2 ランダム型学習の集団では，学習メカニズムが貧弱で，可塑性のメリットとコストのバランスがうまく取りきれないからであると考えられる．

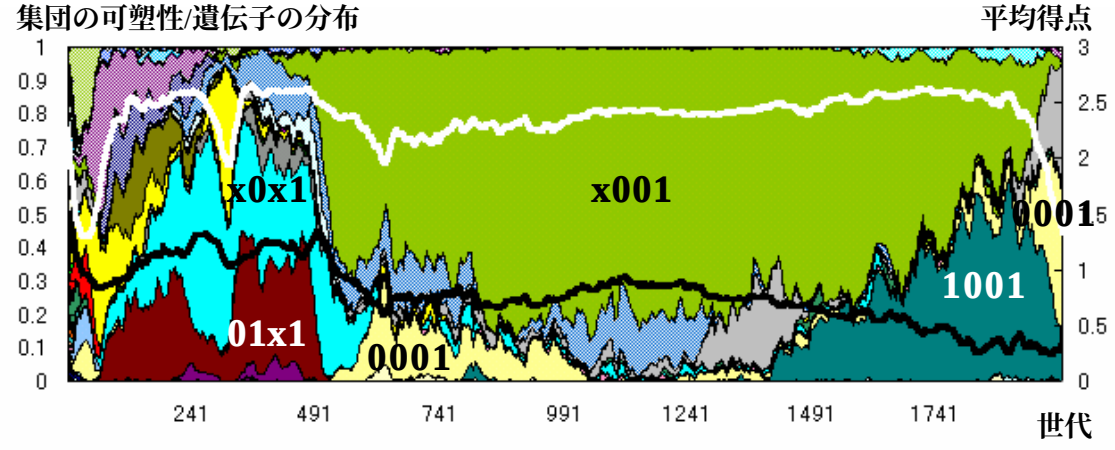

図 5：記憶長 2 ランダム学習での実験結果

5.3 記憶長 2 メタ・パブロフ学習

次に，記憶長 2（初期集団における GS，GL 各遺伝子の値はランダム）メタ・パブロフ学習の集団において同様のパラメータを用いて進化実験を行った.

実験結果の一例として図 6, 7 を示す. 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 250 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.

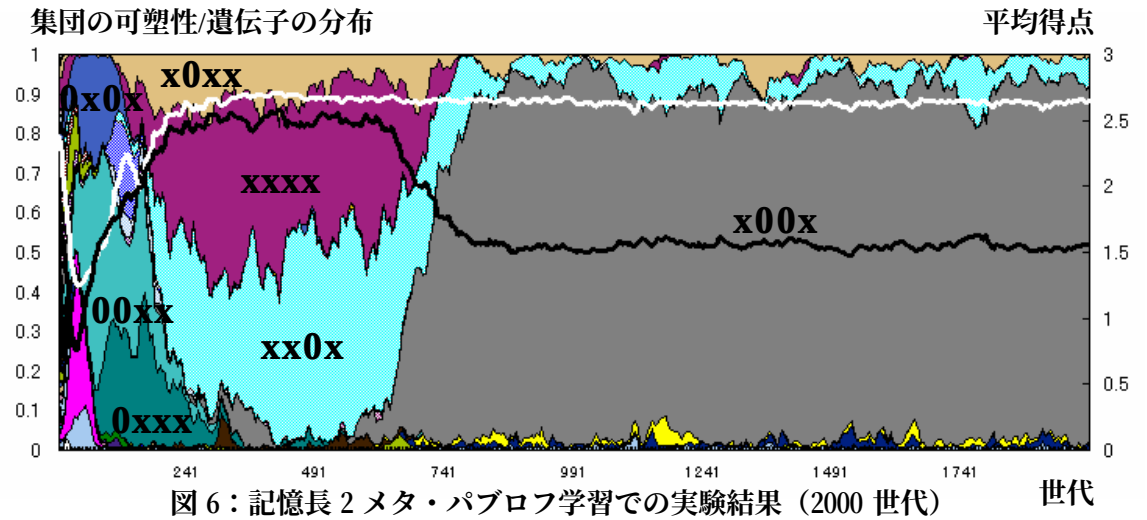

図 6：記憶長 2 メタ・パブロフ学習での実験結果（2000 世代）

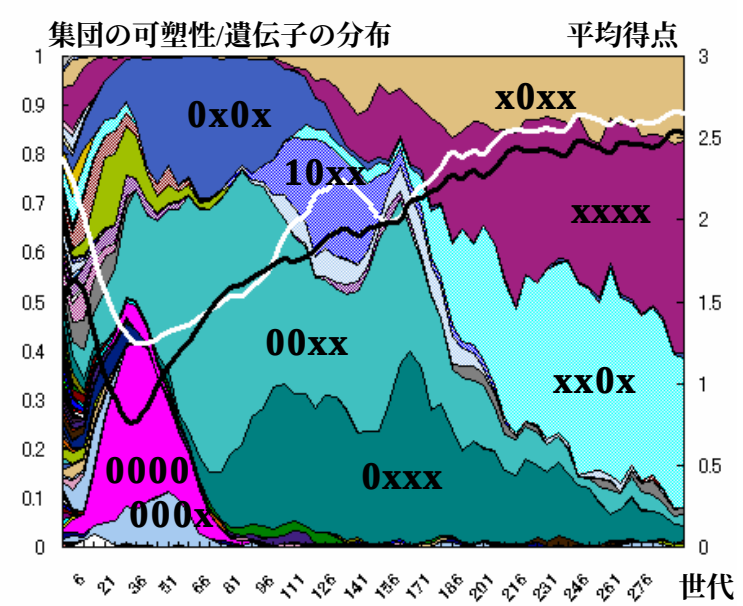

図 7：記憶長 2 メタ・パブロフ学習での実験結果（300 世代）

5.4 戦略の推移と Baldwin 効果

これまでの結果から，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，可塑性は戦略集団の進化にどのような影響を与えたのだろうか．5.3 節の記憶長 2 メタ・パブロフ学習の実験における戦略の推移と Baldwin 効果の関係について詳細に解析する．

集団の可塑性と平均得点の相関に注目して進化の過程を観察すると，Baldwin 効果の 2 つの段階をはっきりと把握することができる．図 8 は 10 回の試行における，集団の可塑性と平均得点の相関の軌跡を表したものである．初期集団の状態から，一旦平均得点と集団の可塑性が低い状態へと進化した後，共に上昇する方向（グラフ右上）へと進化しているのがわかる．これが Baldwin 効果の第 1 段階である．その後，軌跡はまっすぐ左へと向きを変え，平均得点を維持したまま，集団の可塑性のみが低下しているのがわかる．これが第 2 段階である．

この相関図の軌跡にあわせて，出現した戦略を大まかに分類すると，図 9 のようになる．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が多く含まれるため，相関図左下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．左下の裏切り的な戦略群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，右下のような戦略が

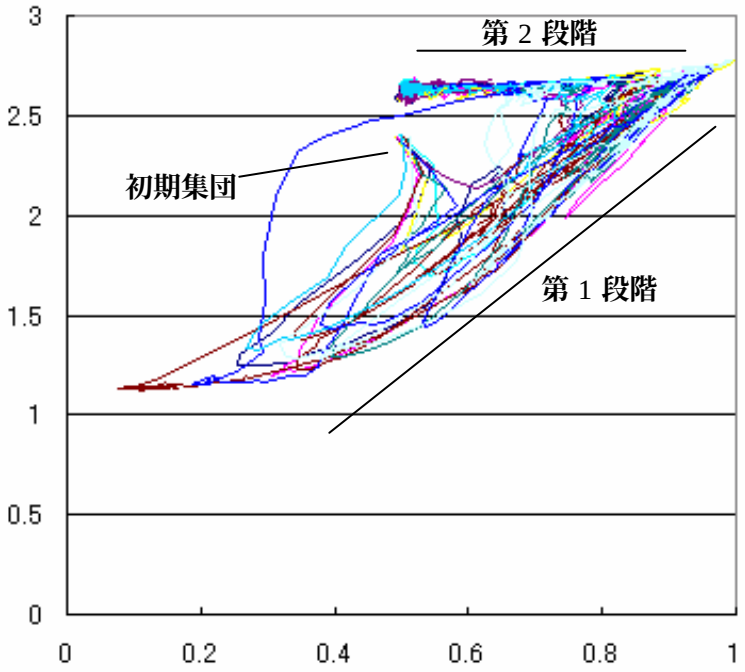

図 8：進化の過程における集団の可塑性と平均得点の相関

徐々に集団中に広まる．これらの戦略は左下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，相関図右下のような戦略が増加し，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，右下の戦略と比べてより協調的であることで高い得点を得る右上のような，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[x00x]型戦略よりも右上の戦略群の方へと進化する傾向が高いのは興味深い．左下から右下，右上の戦略群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，[x00x]の ESS 的な性質（6.1 節で後述）が右下の戦略群との対戦で両者の得る得点を，右上の戦略群と比べて下げてしまっているので，右上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，右上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

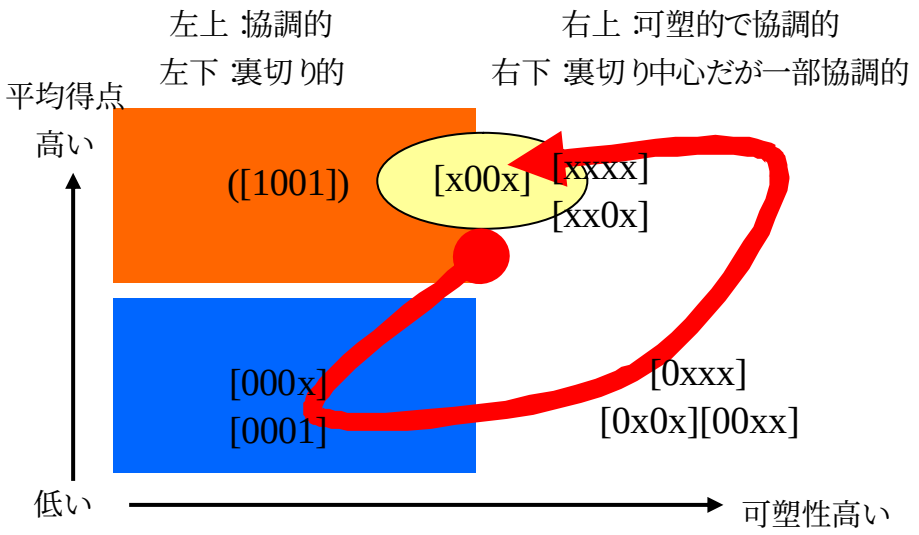

図 9：集団の可塑性・平均得点と進化の過程で出現した戦略の分布

は「協調の維持に必要最小限の可塑性を持った戦略」[x00x]型戦略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(6.1 節で後述), ここで集団は安定する．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.

6.メタ・パブロフ[x00x]型戦略の解析

5.3 節の結果から，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．そこでこの戦略の安定性や特徴，学習の果たす役割についていくつかの点から解析する．

6.1 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，

$$E(a,a) > E(b,a) \tag{5}$$

または

$$E(a,a) = E(b,a)かつE(a,b) > E(b,b) \tag{6}$$

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．

メタ・パブロフ[x00x]型戦略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 10 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（5）式の左辺と右辺の差である．したがって，[x00x]が ESS であるには同種同士の対戦を除いて縦軸の値がすべて 0 より大であれば良いが，この図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．

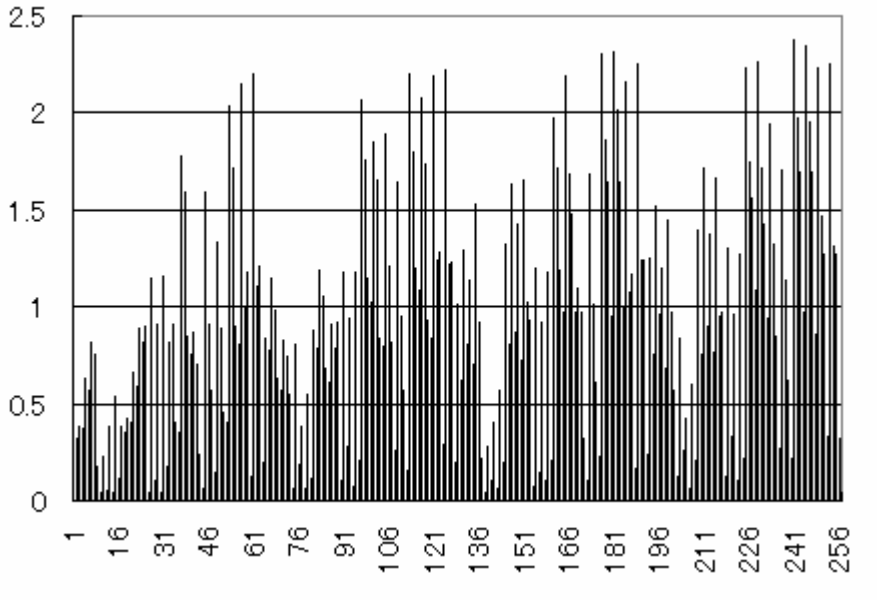

図 10：メタ・パブロフ[x00x]型戦略と他の記憶長 2 の戦略との対戦成績

6.2 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 11 に[x00x]型戦略の状態遷移の様子を示す．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．

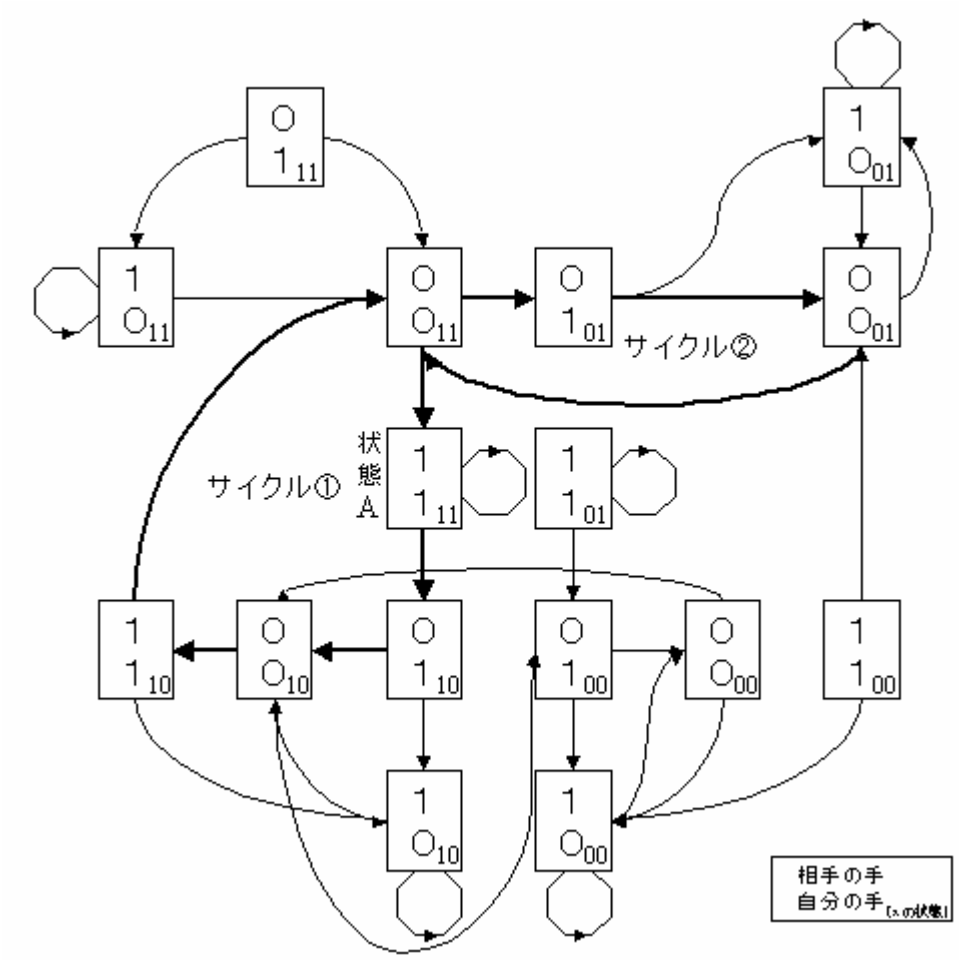

図 11 メタ・パブロフ[x00x]型戦略の状態遷移図

6.3 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，進入させないという点で有効であると考えられる．図 11 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 回の裏切りあいの後 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．

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

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 11 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．

表 4：[x001]と[x00x]の対戦例（下線はノイズ）

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

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.

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…

7. より一般化した進化実験

5 章では，基本的な記憶長 2 のモデルにおいて進化実験を行った．本章では，5 章のモデルに対し記憶長と学習行列の 2 つの点から一般化を行ったモデルを採用し，その実験結果について示す．

7.1　記憶長の突然変異を導入した進化

これまでの実験では，各戦略は記憶長が固定されており，遺伝子列であらわすことのできる戦略の種類は限られていた．そこで，記憶長が長くまたは短くなる突然変異を導入することで，理論上戦略が制限されないオープンエンドな環境を設定し，進化実験を行った．

具体的には，遺伝的オペレーションにより次世代の個体を生成する際，各個体につき確率 *1/3000* で GS,GL 各遺伝子列の長さを 2 倍（各遺伝子列について，遺伝子列を 2 つならべたものを新たな遺伝子とする）または 1/2 倍（各遺伝子列について，遺伝子列を 2 等分し，そのうちどちらかを新たな遺伝子列とする）するものとした（例：2 倍[1101]→[11011101]，1/2 倍[1101]→[11]または[01]）．なお，この方法は Lindgren らのモデル[Lindgren 91]に基づいているが，記憶長が増加する方への突然変異において，彼らのモデルの場合は戦略自体の挙動は変化しないが，この研究においては学習遺伝子列の影響により戦略自体の挙動も変化する点で異なるものである．

初期集団を，GS はランダム，GL=[00]とした記憶長 1 のメタ・パブロフ学習を行う集団として行った典型的な結果を図 12 に示す．試行によってばらつきはあるが，数百世代にわたって[0101]（しっぺ返し戦略）など記憶長 1 の戦略が交互に集団中を

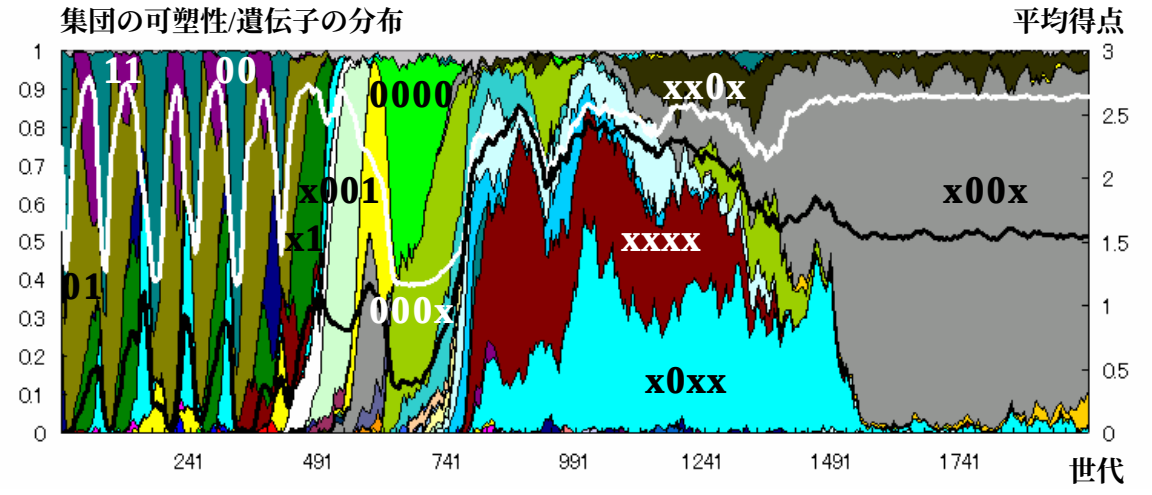

図 12：記憶長の突然変異を導入した場合の実験結果

占めたのち，記憶長 1 の集団から記憶長 2 の集団へと進化した．その後，試行の 50% で 5.3 節での過程と同様な進化が見られ，それ以外では Baldwin 効果（集団の可塑性の増加と減少）は見られないまま，直接[x00x]型の集団へと進化した．前者の場合は可塑的で協調的な戦略（[xxxx]など）が現れる前に，裏切り的な集団が集団中を大きく占め，平均得点を低下させている場合が多く，後者の場合，[x00x]型の戦略が現れる前に，すでに協調的な集団になっている場合が多く見られた．これは，集団の可塑性の上昇が，裏切り的な集団から協調的な集団へのシフトとその後の[x00x]型戦略へのスムースな進化に貢献しているのを支持するものと考えられる．

また，ごくわずかな場合において，メタ・パブロフ[x00x]型戦略に収束せず，記憶長３以上の戦略が多種類集団中を占めることが確認されたが，一旦[x00x]型戦略の集団に進化した後は，それ以上記憶長の長い集団へと進化することはほとんど見られなかった．このことは，メタ・パブロフ学習を伴う場合には，記憶長 2 の戦略[x00x]が十分安定であり，それ以上記憶長の長い戦略は集団の安定のために必要とされなかったことを示している．

Tuomas らは，強化学習を行うマルチエージェント系において，各エージェント自体の学習が他のエージェントにとって不安定な環境を作りだしてしまい学習がうまくいかず，結果として集団全体の利益を下げてしまう場合があることを指摘している[Tuomas and Robert 96]．メタ・パブロフ学習は強化学習の中で最もシンプルな学習方式の一つとして考えられるが，この結果は記憶長が長くなり学習が複雑になるほどよいというわけではないことを示している．従って，マルチエージェント系における強化学習においては，自らが学習することで他のエージェントに与える影響の大きさについても考慮する必要があると考えられる．実際，GS=[0000]GL=[1001]で表される[x00x]型戦略が記憶長の突然変異を起こしてできた個体（GS=[00000000]GL=[10011001]）同士が対戦を行うと，一方の学習による挙動の変化が他方の学習を邪魔して協調し合えない状態が続くことが観察された．

7.2 学習行列を遺伝子として取り込んだ進化

これまでの学習行列を用いた実験では，行列の値としてメタ・パブロフ学習行列を定義し，前もって与えてきた．しかし，これ以外の学習行列を持つ戦略がどのように振る舞うかは興味深い点である．そこで学習行列を表す遺伝子列 GT を新たに 3 つ目の遺伝子列として定義する．各個体はそれぞれ GT を持ち，GT の表す学習行列を参照して学習を行うものとして進化実験を行った．GT は具体的に表 6 のように定義される（例：メタ・パブロフ学習行列:GT=[1001]）．

表 6：学習遺伝子列 GT の定義

相手の手（→） 自分の手（↓）	C	D
C	(CC)	(CD)
D	(DC)	(DD)

GT=[(DD)(DC)(CD)(CC)]

以上のような変更をモデルに加え，進化実験を 4000 世代に渡って行った．初期集団は遺伝子をランダムに決定した 100 個体をそれぞれ 10 個体ずつ用意するものとした．進化実験の結果を図 13,14 に示す．図中の####:****はこれまでの遺伝子表記における記述が####，GT が****の個体であることを示す．なお，学習行列の進化を導入することにより，集団の可塑性は集団の学習に対する依存度を反映しにくくなるため，考察には含めない．

60 回試行を行ったところ，そのうち 29 回で [x00x]GT=[1001]すなわちメタ・パブロフ学習行列を用いた戦略が集団中を占めた（図 13）．また，そのうち 3 回で [x001]GT=[1000]が集団中を占めた（図 14）．残りの 28 回ではひとつの戦略には収束しない不安定な状態であった．学習行列を限定しなくても最終的にメタ・パブロフ [x00x]型戦略が頻繁に集団中を占めたことから，メタ・パブロフ学習行列は学習則として妥当なもののひとつであると言える．

全 60 試行について 4000 世代の時点で集団の 5%以上を占めた個体についてその割合を多い順に示したものが図 15 である．[x00x]GT=[1001]が半分弱を占めているのがわかる．この中で[xxxx]GT=[0001]，[xxx1]GT=[0001]は収束することはなかったものの[x00x]GT=[1001]に次いで高い割合を占めている．これらの戦略は集団中に頻繁に現れるが，安定ではないので他の戦略の侵入を許してしまうからであり，完全に収束した[x00x]GT=[1001]や[x001]GT=[1000]とは異なる．

[x001]GT=[1000]型戦略は Boerlijst らが提案した pPAVLOV 戦略[Boerlijst, Nowak and Sigmund 97]とほぼ同等の振る舞いをすることがわかっている．図 16 は両戦略を

状態遷移図で表したものである．協調の回復過程において両者は非常に類似しており，どちらも「2 回の裏切りの後のみ協調に転じる」性質を持っている．今回のモデルのようなシンプルな進化の枠組みにおいて，このような手作りの戦略が現れたのは興味深い．また彼らは，ジレンマゲームにおける戦略において，状況に応じてタグ付けを行う戦略の重要性を主張し，PAVLOV 戦略に代わるロバストな戦略として pPAVLOV 戦略を提案しているが，それ以上にメタ・パブロフ[x00x]型戦略が頻繁に集団中を占めたという結果は，タグ付けの方法の一つとして，メタ・パブロフ学習行列による学習のような強化学習的なタグ付けが有力であることをしていると言える．

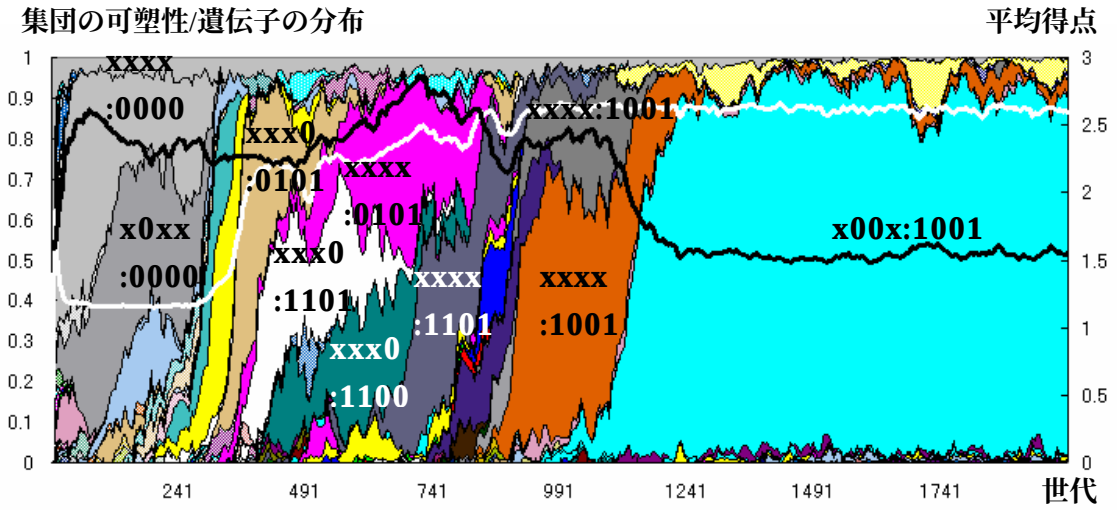

図 13：学習行列を遺伝子に取り込んだ実験結果
（メタ・パブロフ[x00x]型戦略に収束した例）

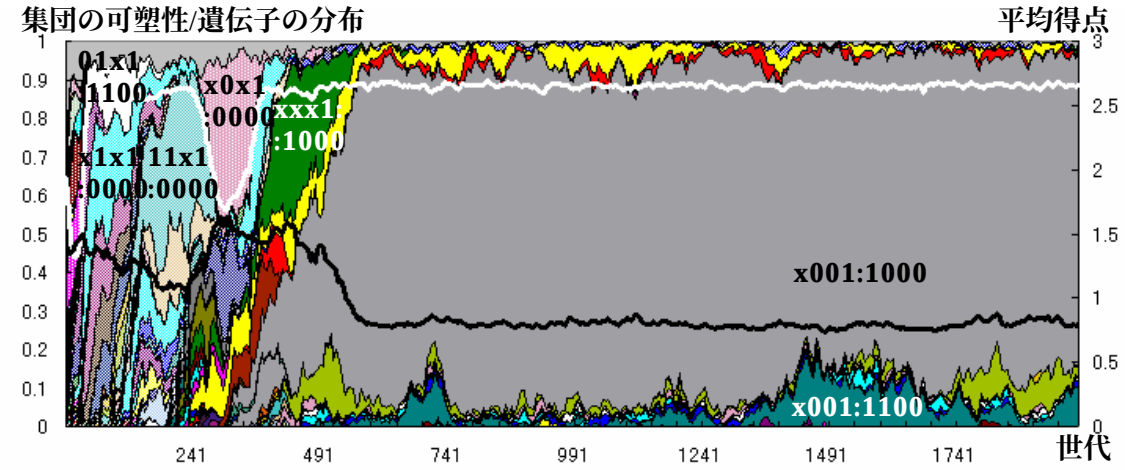

図 14：学習行列を遺伝子に取り込んだ実験結果
（[x001],GT=[1000]に収束した例）

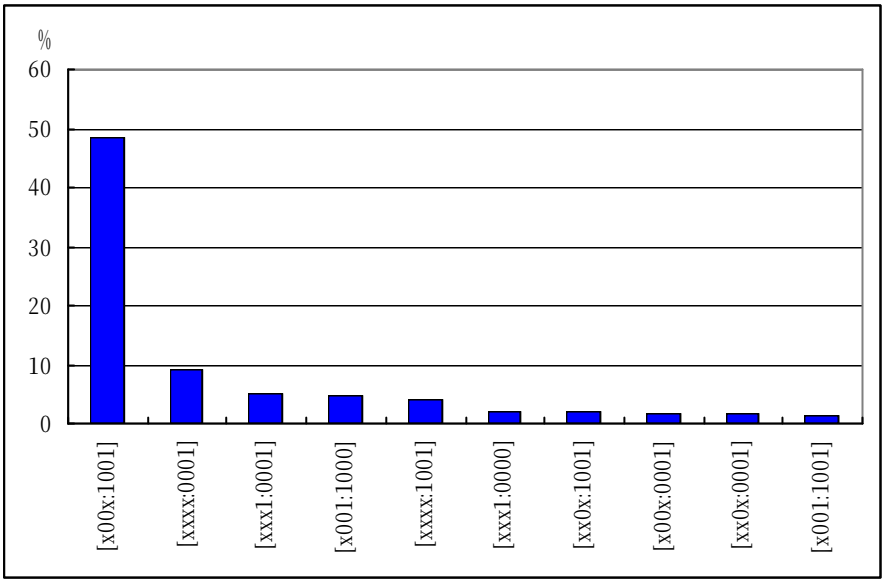

図 15：4000 世代目において各戦略が集団中を占めた割合（60 試行）

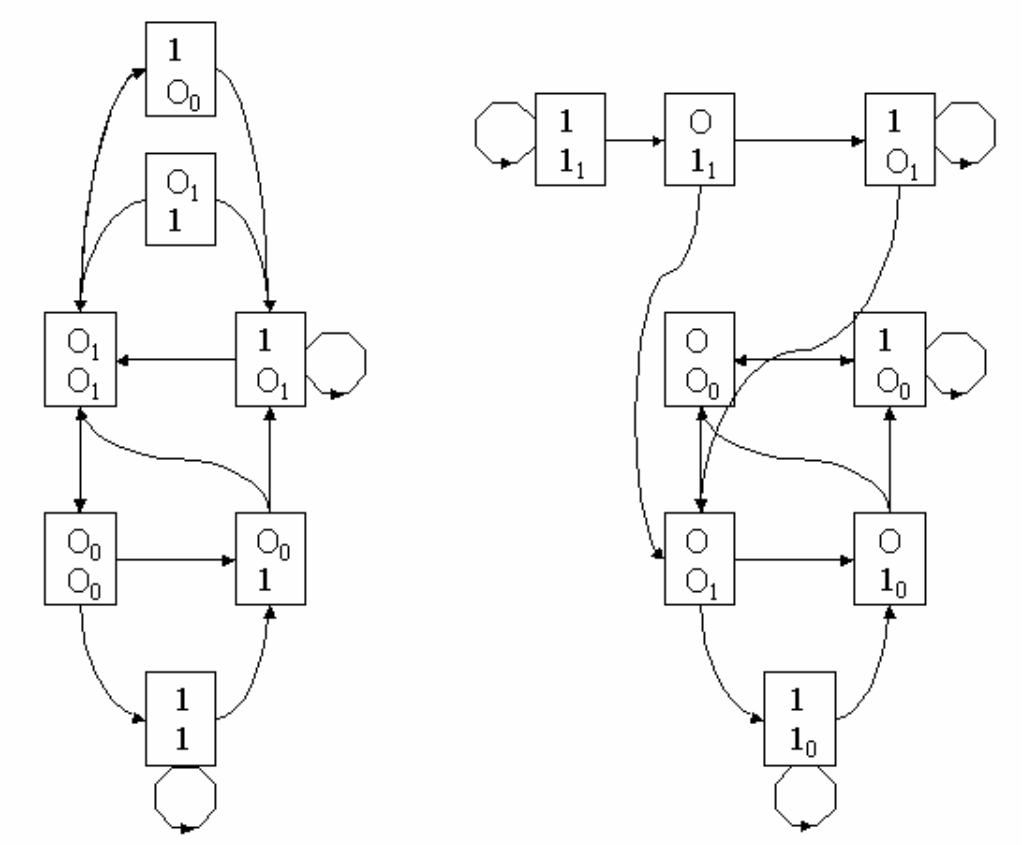

図 16：pPAVLOV（左）と[x001]GT=[1000]（右）の状態遷移図

8. 結論

8.1 まとめ

本研究では，動的な環境における進化と学習の相互作用について知見を得るために，繰り返し囚人のジレンマゲームにおける戦略にメタ・パブロフ学習に基づく表現型の可塑性を導入して進化実験を行い，その結果について考察してきた．

初めに，基本的な実験として記憶長 2 の戦略集団を用いて実験を行った．ランダム型学習を用いる集団では，ある特定の戦略に完全に収束することは無かったが，メタ・パブロフ学習を行う集団においては Baldwin 効果の 2 つの段階が働いた結果 [x00x]型の戦略が集団中を占めた．

Baldwin 効果が確認された過程において，可塑性が戦略の進化に与える影響を具体的に解析した．その結果，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，Baldwin 効果の第 1 段階として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，第 2 段階として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．本研究においては学習の際の明示的なコスト（およびメリット）を与えないとしたが，解析の結果，このモデルにおける学習のメリットは裏切り的な集団から協調的な集団へ進化するための原動力であり，またコストは他の戦略に侵略される可能性として働いていることがわかった．このメリットとコストは，どちらも動的環境に特有のものであり，動的環境における Baldwin 効果の存在を示すものである．

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]型戦略は記憶長 2 の集団において ESS の条件を満たすなど，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．また，記憶長の突然変異や学習行列の進化を導入し，より一般化したモデルにおいても，多くの場合において[x00x]型戦略が集団中を占めることがわかった．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．

8.2 今後の展開

本研究はジレンマゲームにおける戦略の進化という抽象的な環境での議論であるが，それゆえに多様な発展が可能であると考えられる．その可能性としていくつかの方向性を挙げることが出来る．ひとつは，現在のモデルについての解析および拡張をさらに行い，進化と学習の相互作用に関する考察を深めることである．たとえ

ば，ジレンマゲームの利得行列を変動させて，環境自体の変動を導入するのは興味深い．また，より長い記憶長を用いた実験や，別の学習方式を用いた実験なども考えられる．

もうひとつは，このような動的環境における進化と学習の相互作用の仕組みの工学的応用である．本研究で扱ったメタ・パブロフ学習を最もシンプルな強化学習メカニズムの一つと見なすと，集団全体はマルチエージェント強化学習系として捉えることができる．近年，強化学習に関する研究が盛んに行われている中で，マルチエージェント系における強化学習では，個々のエージェントの学習が互いの環境を不安定にし，学習を困難にさせてしまうことが問題となっている．このような問題に対し，本研究で用いたような集団全体の学習への依存度が量的に把握でき，また依存度が進化の過程で調節されるようなモデリングは有効であると考えられる．現在，クラシファイアシステムを用いた強化学習エージェントを用いて，応用の可能性について検討している．

また，より生物学的な側面への発展のひとつとして，鳥の鳴き声の性選択による進化に関する研究が挙げられる[Todd 96]．鳥の鳴き声は，幼少期における成鳥からの学習に依存する要素と，先天的に決定されている要素に分けられる．その比重は種や環境によってさまざまであり，これは進化の過程でうまく調節された結果であると推測できる．この過程において，本研究で示したような進化と学習の相互作用が起こっている可能性は十分考えられる．モデルを構築し，様々な環境で実験を行うことで，鳥の鳴き声の進化に関する新たな知見が得られると考えている．

最後に，表現型の可塑性に関する議論を突き詰めて行くと，文化の進化に関する議論に発展して行くと考えられる．集団が遺伝的な拘束から乖離しつつ，何らかの秩序を構成していく過程について，このような集団の可塑性に注目したモデリングをもとに議論できる可能性があると考えられる．

謝辞

本研究を進めるにあたり，貴重な時間を割いて御教授，御助言を頂きました，名古屋大学大学院人間情報学研究科助教授 有田隆也先生に心より感謝申し上げます．

また，公私共に様々な面で御支援下さった，有田研究室の諸氏に感謝致します．

最後に，これまで暖かく見守って下さった両親に心より感謝申し上げます．

参考文献

[Ackley and Littman 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 2000] 有田隆也: ***人工生命***, 科学技術出版 (2000).

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

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

[Boerlijst, Nowak and Sigmund 97] Boerlijst, M. C., Nowak, M. A. and Sigmund, K.: The Logic of Contrition, *Journal of Theoretical Biology*, Vol. 185, pp. 281-293 (1997).

[Harvey 96] Harvey, I.: Is There Another New Factor in Evolution?, *Evolutionary Computation*, Vol. 4, No.3, pp. 313-329 (1996).

[Hinton and Nowlan 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 and Sigmund 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).

[鈴木 99] 鈴木麗璽: 囚人のジレンマゲームにおける Baldwin 効果, *ALIREN（人工生命研究会）+ALIST（在京若手研究者による人工生命自主セミナー）合同研究会ポジションペーパー* (1999).

[鈴木, 有田 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, 有田 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).

[鈴木, 有田 99c] 鈴木麗璽, 有田隆也: 学習が進化の過程に与える影響に関するシミュレーション解析, *第 9 回数理生物学シンポジウム予稿集（数理生物学懇談会ニュースレター）*, p. 9 (1999).

[鈴木, 有田 2000] 鈴木麗璽, 有田隆也: 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 -, *人工知能学会誌*, Vol. 15, No. 3 (採録決

定).

[Suzuki and Arita 2000] Suzuki, R. and Arita, T.: How learning Can Affect the Course of Evolution in Dynamic environments, *Proceedings of the Fifth International Symposium on Artificial Life and Robotics*, pp. 260-263 (2000).

[**佐々木, 所** 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).

[Todd 96] Todd, P. M.: Sexual Selection and Evolution of Learning, *Adaptive Individuals in Evolving Populations*, Vol. 26, pp. 365-393 (1996).

[Tuomas and Robert 96] Tuomas, W. S., Robert, H. C.: Multiagent reinforcement learning in the Iterated Prisoner's Dilemma, *Biosystems* 37, pp. 147-166 (1996).

[Turney, Whitley and Anderson 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).