

# 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

鈴木 麗璽 有田 隆也

名古屋大学 大学院人間情報学研究科  
Graduate School of Human Informatics, Nagoya University  
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan  
{reiji, ari}@info.human.nagoya-u.ac.jp

## 講演要旨

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。

これまで、Baldwin 効果に関する進化実験による研究は Hinton と Nowlan による先駆的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。

本研究では、特に個体間の相互作用にのみに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、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。

そこで、戦略集団の各個体を、囚人のジレンマゲームの戦略において決定論的戦略を表す遺伝子列 GS と、GS の表現型が可塑的であるかどうかを示す遺伝子列 GL によって定義し、我々の提案するメタ・パブプロフ学習行列に従って表現型を変更するものとして、進化実験を行った。

その結果、Baldwin 効果の第 1 段階として、集団全体における可塑性の増加を伴った、裏切りのない集団から協調的な集団へのシフト、第 2 段階として協調関係を維持しながらも協調集団の維持に必要な最小限の可塑性を持ったメタ・パブプロフ [x00x] 型戦略への収束が確認された。

# 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

鈴木 麗壘 有田 隆也

名古屋大学 大学院人間情報学研究科  
Graduate School of Human Informatics, Nagoya University  
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan  
{reiji, ari}@info.human.nagoya-u.ac.jp

## 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 効果がある。この現象は、ラマルクの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。この効果は、一般には以下の 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。

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。

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、Baldwin 効果に関する進化実験による研究は Hinton と Nowlan による先駆的実験[1]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、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

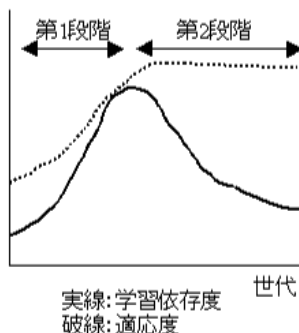


図 1：適応度と学習依存度から見た Baldwin 効果

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、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。

## 2. 表現型の可塑性を導入したくり返し囚人のジレンマゲームの戦略の進化モデル

この研究では、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、くり返し囚人のジレンマゲームの進化実験を取り上げ、我々の提案するメタ・パブプロフ学習を戦略に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。構築したモデルは次のとおりである。

くり返し囚人のジレンマゲームは、社会的集団に生じるジレンマ状態をシンプルに抽象化したモデルであるが、この研究では以下の利得行列を用いる。

表 1：囚人のジレンマゲームの利得行列

相手の手(→)	協調	裏切り
自分の手(↓)	(C)	(D)
協調 (C)	(3, 3)	(0, 5)
裏切り (D)	(5, 0)	(1, 1)

(自分の得点, 相手の得点)

集団中の各個体の持つジレンマゲームにおける戦略は、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される。戦略遺伝子列は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。

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

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  $A_k$  とすると、記憶長  $m$  の戦略は、

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

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  $A_k$  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、学習遺伝子列を

$$GL = [L_0, L_1, \dots, L_{n-1}] \quad (3)$$

と定義する。

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、メタ・パブロフ学習行列(表2)に基づいて表現型を変更(学習)する。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。
- 次回対戦以降、新たな表現型を参照し手を決定する。

表2: メタ・パブロフ学習行列

相手の手(→)	C	D
自分の手(↓)		
C	C	D
D	D	C

以上のような戦略を持つ集団について、ノイズありのくり返し囚人のジレンマゲームの総当たり戦を行い、その合計得点を各個体の適応度としたルーレット選択により次世代の集団を作り出す操作をくり返し、集団を進化させる。その際、遺伝子の一点突然変異を導入する。

### 3. 実験結果と考察

記憶長2(初期集団はランダム)の集団において、パラメータとして突然変異率0.001, 個体数1000, ノイズ率2% 未来係数0.99, 世代数2000を用いて進化実験を行った。その結果、試行の約75%で図2のような傾向を持つ進化が確認され

た。なお、集団中を占める戦略個体は、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してある(e.g. GS=[1000], GL=[1001] [x00x])。

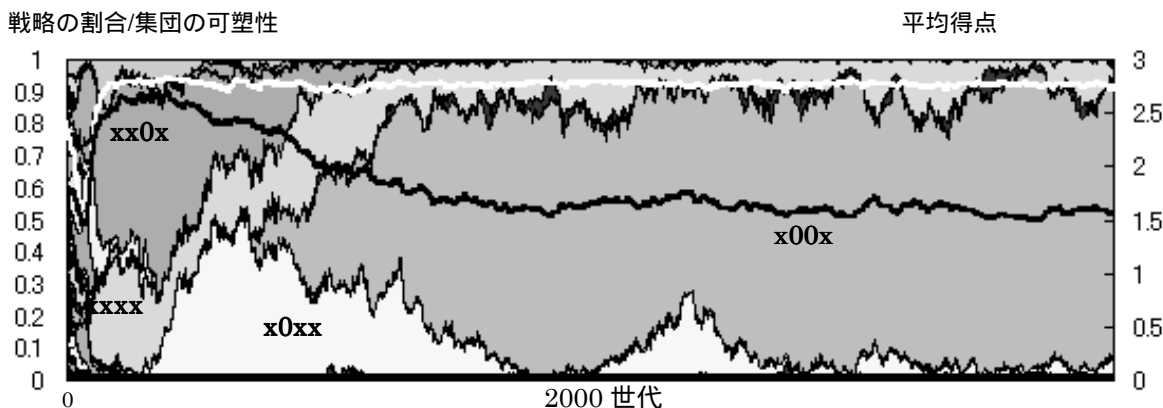
はじめの約60世代では、裏切りの戦略が集団中を占め、平均得点が低下した。その後、集団の可塑性(学習遺伝子列中に占める"1"のビットの割合)の増加とともに平均得点は上昇し、約120世代までに高い平均得点を維持する協調的な集団へと進化した。これは、可塑性が裏切りの集団から協調的な集団へのシフトに有利な方向へと働いたことを示し、Baldwin効果の第1段階と考えられる。

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、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、十分協調関係が築かれた集団においては、ノイズによって可塑性がコストとして働くため、集団を維持するために最低限必要な可塑性のみが選択されたことを示しており、Baldwin効果の第2段階と考えられる。

現在は、メタ・パブロフ[x00x]型戦略の解析[2, 3], 記憶長を固定しないオープンエンドな進化実験や、学習行列自体も遺伝子として表現し、進化させる実験などを行っている。

### 参考文献

- [1] G.E. Hinton and S.J. Nowlan(1987): How Learning Can Guide Evolution, Complex Systems, Vol.1, pp.495-502.
- [2] 鈴木 麗璽, 有田 隆也(1999): メタ・パブロフ: 進化と学習による適応を自動調節する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.
- [3] Suzuki R. and Arita T. (2000): How Learning Can Affect the Course of Evolution in Dynamic Environments: Fifth International Symposium on Artificial Life and Robotics (to appear).



(平均得点: 白実線、集団の可塑性: 黒実線)

図2: 記憶長2での実験結果