

GAによる囚人のジレンマ実験

—ラマルク的手法とダーウィンの手法—

安 田 俊 一

0 はじめに

Robert Axelrod が “The Evolution of Cooperation” を発表して 20 年近くになる。この本とそれに先立つ囚人のジレンマに関する一連のコンピュータトーナメントについての論文 (Axelrod (1980a, 1980b)) は、ゲーム理論だけでなく、社会学・生物学・政治学など広範な分野に大きな反響を巻き起こした¹⁾

Axelrod の実験は一般的に大変高い評価を得た²⁾。彼はこの実験から「Tit for tat (TFT)」戦略³⁾が繰り返し囚人のジレンマ (Repeated Prisoner's Dilemma, 以下 RPD) において非常に有効な戦略であると主張した。この主張に対して Boyd & Lorberbaum (1987) が有限繰り返し RPD には純戦略の範囲では進化的に安定な戦略はないことを証明したのをはじめとして、RPD における TFT 戦略の安定性に関する検討はほぼ終わっている。その後、研究の流れは無限 RPD で出現する「協力」解や解のクラスを検討する方向へ進んでいった。

本稿では無限 RPD ではなく有限 RPD の範囲において、従来の論者があまり取り上げていない Axelrod の遺伝的アルゴリズム (以下, GA) を使った実験

1) このコンピュータトーナメント以降の Axelrod による同テーマに関する考察は Axelrod (1984) にまとめられている。

2) この実験に関する問題点も数多く指摘されている。それらについては Hoffmann (2000) による簡潔なまとめを参照。

3) 「しっぺ返し」戦略。最初に「協力」し、あとは前回の相手の手をまねるというもの (後述)。

(Axelrod (1987)) についての問題点を検討する。

現実問題として生物や組織が囚人のジレンマ的な状況におかれた際には無限 RPD ゲームを行うと考えるよりは終了回数がプレイヤーに認識されていない有限 RPD を行うと考えた方が自然であり、またプレイヤーはそれほど聡明ではないので、有限オートマタを用いたメタゲームを解いて協力解に到達するかどうかを考察する (例えば Linster (1992)) ことも重要だが、一方で、GA のように単純なシステムの中から TFT 戦略が発生してくるかどうか、また「協力の進化」が起こるのかどうかを問題にすることも依然として重要であるからだ。

ここでは Axelrod の GA 実験における以下の3つの問題点を検討する。

最初に GA のコーディングに関する問題を取り上げる。Axelrod 実験では GA のコーディングの際に「ラマルクの進化」ともいえる手法を採用している。これは一般に生物行動の進化を検討する際には根拠が薄弱である。Axelrod は過去のゲームの記憶を遺伝子型 (Genotype) に含めているが、この発想は進化の方法としては「獲得形質の遺伝」に通じるものであり「ラマルク的手法」と呼べるであろう。しかし厳密な意味で遺伝子型を取り上げるのであれば、過去の対戦記憶が遺伝子型の一部として遺伝するという手続きは問題が多い。本稿では「ダーウィンの手法」に基づくコーディングを行う。

つづいて Axelrod の主張を検証する。彼は GA を使った実験で「TFT によく似た戦略が進化してきた」と主張するが、その際に GA によって生み出されてきた遺伝子配列パターンが TFT であるとする基準が曖昧である。ここでは GA の枠組みの中でのそれを特定し、その基準に照らして戦略の進化を検討する。

最後に TFT 戦略の影響力をシミュレーションによって考察する。これは GA における個体のように本質的には無能な場合にでも、もしなんらかの形で TFT 戦略タイプが与えられた場合にそれを学ぶことができるかどうか、という問題の考察である。本稿で行うように、GA 実験において TFT をゲームの環境として取り扱ったシミュレーションはこれまでにはないと思われる。⁴⁾シミュ

レーションの結果から、個体が TFT 戦略者とゲームを行う場合には集団全体が TFT 戦略を採っているかのようなふるまいをみせるようになるし、またその結果協力の進化が起りやすくなることが明らかになる。

以下では Axelrod の一連の実験を概説した後、本稿でもちいる「ダーウィンの手法」に基づくシミュレーションモデルを概説する。つづいてそのシミュレーションモデルに基づき実験結果を述べる。

1 Axelrod 実験 1 : コンピュータトーナメント

はじめに Axelrod の最初の実験である繰り返し囚人のジレンマに関するコンピュータトーナメントに関して概説し、そこから彼が引き出した結論を整理する。

1.1 囚人のジレンマ概要

囚人のジレンマゲームは以下の利得行列に基づくゲームである。各プレイヤーはそれぞれ「C : 協力 (cooperation)」、「D : 裏切り (defect)」という選択肢を持つ⁵⁾

	C	D
C	(R, R)	(S, T)
D	(T, S)	(P, P)

ただし、 $T > R > P > S$, $R > (T + S)/2$.

この場合、双方のプレイヤーにとって D 戦略は支配戦略であり、その結果両

4) 集団中にいくつかの戦略を混合させた研究には Nowak & Sigmund (1992) がある。これは Axelrod (1980b) の後半の実験をより生物に近づけた仮定で行ったようなシミュレーションであり本稿の GA 実験とは全く異なるものである。

5) 「協力」「裏切り」と名付けられているのはこのゲームの代表的な解釈が共犯者に「協力」して沈黙を守るか、共犯者を「裏切っ」て自首してしまうか、という状況を想定しているからである。だから協力は「沈黙」、裏切りは「自首」とよばれることも多い。

者とも相手を裏切ってしまう。つまり、 (D, D) はこのゲームの唯一の Nash 均衡点になっている。

しかし $R > P$ なのだから両者ともに C 戦略をとれば両者とも利得は増加する。そのことは両者ともにわかっているのだが、相手を裏切ったときの利得の大きさ、あるいは相手に裏切られたときの恐怖から裏切らざるを得ない。このゲームは集団の利益と個人利益が対立する状況で頻繁にあらわれるため長い間研究されてきている。

ジレンマ解決の方法として提案されたのが、1回だけのゲームではなくこのゲームを繰り返して行うことである（繰り返し囚人のジレンマ RPD: Repeated Prisoner's Dilemma）。このとき戦略は単に C か D を出す、ということではなく、どのような順番で C と D を出すか、相手の動きにどう応じるか、というものになる。

RPD の中でもっとも有効な（何回も繰り返した結果累積された利得が大きい）戦略として提案されたのが TFT であった。これは「最初に協力（C）を出した後、相手が協力（C）すれば次回は自分も協力し、相手が裏切れば（D）次回は自分も裏切る」という戦略である。あるいは単純に、「初回は協力し、次回以降は相手の一回前の手をまねる」といってもよい。

1.2 実験概要

Axelrod (1980a, 1980b) は、依頼や公募によって集めた RPD における戦略をコンピュータで対戦させてみた。第1回目のトーナメントでは専門家から提出された14のプログラムとランダムに C と D を出すプログラムの合計15のプログラムでトーナメントを行い、Table. 1 の利得行列に基づく RPD を 200 回対戦させ、その累積利得を比較した。実験においてもっとも高い利得を得たのが TFT であった。

つづいて彼は、公募によって集められた63のプログラム（TFTを含む）同士で対戦させた⁹⁾。結果はふたたび TFT がもっとも優秀であった。

この2回目の実験の時には Axelrod は単にプログラム同士を対戦させただけでなく、対戦で得た平均利得が次のトーナメントでのそのプログラムの比率を決めるような実験を行っている。これは平均利得を次世代に残す子孫の数とみなして集団中でそれぞれの戦略がどのような割合になるかを実験したものである。

ここでも TFT は1,000 世代後に集団中の14.5%を占め（これはごくわずかの差ながらも集団中の最大多数であった）、かつ最大の成長率を示した。

これらの結果から、Axelrod は「TFT は頑健で非常に有効な戦略である」と主張した⁷⁾

	C	D
C	(3, 3)	(0, 5)
D	(5, 0)	(1, 1)

Table. 1 : 利得行列

2 Axelrod 実験 2 : GA

以上2回のトーナメント実験では集められた戦略同士を対戦させただけであるから、この中から新しい戦略が発生してくる要素は当然ない。

そこで Axelrod (1987) は GA を使ってランダムな状態からどのような戦略が発展してくるかを実験した。

この実験は様々なところで参考されているにも関わらず⁸⁾、あまり深く検討さ

6) ただし、この時は確定的な回数ゲームではなく最頻値が200回に近づくように乱数によって回数を変更している。ゲーム回数がわかっているなら、最後のゲームで相手が協力してきた場合、その時に裏切ると確実に利得が増えるからである。

7) もっとも彼は TFT が「最良の」戦略であることは否定している。Axelrod (1980b, 1984)。

8) ほとんどの GA の概説書や入門書はかならずこの実験に言及している（例えば Goldberg (1989), 伊庭齊志 (1994), Mitchell (1996), Michalewicz (1999)）。しかし後述するように、GA からみてこの実験の重要な問題であると考えられる記憶の問題についての説明や解釈はなされていない。

れてきたとは思われない。

ここでは Axelrod の GA 実験をやや詳しく概説し、とくに彼のコーディング手法についての疑問点をあきらかにする。結論を先取りすると、Axelrod は染色体のコーディングにおいて、遺伝子型と表現型を混同している。

2.1 実験の概要

Axelrod は GA のコーディングに際して、「染色体」を「C」と「D」からなる70の長さを持つ文字列配列とした。各文字がゲームにおける選択肢を表している。

最初の6ビット分を「先行する3回のゲームにおける記憶」と定義する（この部分をここでは「記憶領域」と呼ぼう）。第0ビットから「3回前の自分の手、3回前の相手の手、2回前の自分の手…」とならぶ。

「CDDCCC」ならば3回前のゲームでは（自分の手、相手の手）=（C，D），2回前のゲームでは（D，C），1回前には（C，C）が実現したことを示す。

さて、3回分のゲームの記憶は $2^6=64$ 通りの可能な歴史があるから、6ビットの記憶パターンが表現するそれぞれの場合に「次に何を出すか」を定める。

そのために記憶領域6ビットの後ろに64ビットの文字列を準備し（この部分をここでは「戦略領域」と呼ぼう）、記憶領域6ビット中の「C」を「0」「D」を「1」とみなして記憶領域文字列を2進数と考える。

2進数として表された記憶パターンを10進数へ変換し、それが指し示す戦略領域上の場所にある文字を「次の1手」として出すと約束する。

たとえば上の「CDDCCC」は2進数表現では「011000」、10進数に直すと「24」だから、戦略領域の第24ビット目にある文字を出す、と決めるのである。

結果として70ビットの文字列をGAにおける染色体、すなわち「遺伝子型」とすることになる。この場合GAの探索領域は 2^{70} という膨大なものになる。

この染色体を20個ランダムに発生させ、RPDを繰り返したあと、平均利得に応じて「選択」「交叉」「突然変異」というGA過程に入れ、次世代集団を作り

出す。これが実験の概要である⁹⁾

1980年に行われたトーナメント実験の結果から、AxelrodはTFTを除く8つの代表的な戦略を選びだす¹⁰⁾。それらはトーナメント結果をもっともよく表現するものとして回帰分析の結果得られたものである。トーナメントにおいて、ある戦略が得た最終利得はそれら代表的な戦略との対戦結果だけからよく説明された。

Axelrodは8つの代表的戦略それぞれを相手に151回(トーナメントでの平均対戦回数)RPDを繰り返しそれで得た得点の加重平均¹¹⁾を「適合度(Fitness)」としてGA過程を実行した。

2.2 Axelrodの結論

上の8つの代表的戦略から成る固定環境で、GAを50世代くり返す。この実験を40回行った結果からAxelrodは以下の5つの戦略パターンを見いだした。

- (1) CCCCCCに対して「C」。完全な協力
- (2) CCCCCDに対して「D」。直近で相手が裏切ったら次回に裏切り返す
- (3) DCDDCCに対して「C」。「謝罪の受け入れ(Accept an apology)」いったん裏切り、裏切り返されたあとで協力が成立したら、次回も協力する
- (4) CDCCCCに対して「C」。「忘却(Forget)」裏切られても協力がその後に続くと次回も協力する
- (5) DDDDDDに対して「D」。完全な裏切り

Axelrodはこれから「TFTに良く似た戦略パターンが進化した」と結論づけた。

9) 遺伝的アルゴリズムの基本については脚注8)であげた書籍を参考のこと。

10) Axelrod(1980b)では5つの戦略しか得られていないが、1987年の論文では8つになっている。ここでは1987年論文に従う。

11) 加重平均のウェイトは先行のトーナメント実験で得たもの。代表的5つについてはAxelrod(1980b)から得られるが上述のようにGA実験では代表戦略が8つになっているため、実験にAxelrodがどういうウェイトを用いたのかは不明。

この実験は対戦相手を「8つの代表的戦略」としていることからわかるように対戦相手が固定されている。つまりは「固定的環境」での実験である。したがって、この環境自体がTFTを進化させるのに有利に働いたのではないかという疑いが持たれる。

そこで、Axelrodは引き続いて「固定的でない環境」での実験を行った。

対戦相手を8つの代表的な戦略と特定せずに集団の個体同士で対戦させる。これによってゲームの相手は次々に変わり、また世代毎で変わった環境ができあがる。この環境がどう変化していくかを予測することはできない。

こちらの実験ではAxelrodは最終的に得られた戦略パターンを示していない。累積利得の集団平均が世代が進むにつれて増加していくことから「協力の進化」を結論づけただけである。したがってこの変化する環境においてTFT戦略が発生してきたのかどうかについては彼ははっきりとは述べていない¹²⁾

3 問題の所在

以上のAxelrod実験についてはHoffmann (2000)が詳細に問題点を取り上げているが、ここではこれまで誰も指摘していないと思われる問題を取り上げよう。

問題は2つある。コーディングにおける混乱と、TFTの定義に関する曖昧さである。

3.1 コーディング問題

1つめはGAのコーディングにおける「遺伝子型」と「表現型」の問題である。

12) この点、従来の代表的なGAの解説書は誤解を招くような記載を行っている。TFTに似た戦略が進化してきたのはあくまで「代表的な8つの戦略」が固定された環境においてであって、集団内でのランダム対戦という変化する環境においてではない。

Axelrod の手法に従えば染色体の先頭 6 ビットが「記憶」である。記憶は実際に出された手の記録であるから、この部分は表現型¹³⁾であることになる。

ところがこの部分もまた染色体を構成している部分になるので、交叉過程をつうじてこの部分のパターンは染色体中の後半 64 ビットの戦略領域と一部置き換わることになる。戦略領域の部分は遺伝子型¹⁴⁾と考えられるから、この場合、次世代へ伝えられた染色体中には交叉を通じて前世代の記憶領域のパターンに置き換えられた部分が混じっている。その部分は新世代がゲームに入るときに何を意味するのかがはっきりしない。

Axelrod や彼に続く論者のコーディングでは遺伝子型としてバイナリ文字列 (“C”, “D”) を使い、その一つの文字が表現型 (対戦における選択肢) ともなっているため混同されているが、記憶領域が実際には 2 進数として戦略領域から遺伝子 (同時に表現型) を選択することを考えると、記憶領域と戦略領域を同列に取り扱ってはならないのである。

このことは以下のケースを考えるとすぐにわかる。

記憶領域は最初から 2 進数としてコーディングできるのであるから、例えば記憶領域を「0, 1」の整数列としてコーディングした場合、この部分と後ろにある長さ 64 の文字列配列を含めた全体を交叉過程に入れてしまうと、「C, D」の文字列の中に異質な文字 (0.1) が混じることになって次世代の染色体は意味をなさなくなる。

Axelrod のコーディングが有効なのは、遺伝子型と表現型が同じ文字から構成されていることに加えて、過去の対戦記憶を 2 進数に読み替える手続きを背後に隠しているからである。すなわち、コーディング自身の特殊性に基づくものであって、実際には戦略領域が示す遺伝子型の進化に、表現型という別次元

13) 個体の特徴を表現するもの。例えば「尾の長さ」「色の黒さ」などは表現型である。自然選択は表現型をもとに行われる。RPD では「C」「D」という実際に出す手によってゲームの結果が決まり、個体の適合度 (利得) が決まるので記憶は表現型である。

14) どういう表現型を個体に与えるかを決めているのが遺伝子型である。ここでは記憶パターンによって示される個所の文字 (C or D) を「次の一手」として個体に採らせるので戦略領域は遺伝子型である。

のものを紛れ込ませていることになる。

この場合、記憶自身が「遺伝子」として次世代に残されていくのであるから「獲得形質が直接に遺伝する」という「ラマルクの」な手法に基づいている。そう考えるのであれば記憶が遺伝する事自身はいっこうに構わないが、その場合でも記憶領域と戦略領域を同じレベルにおいて操作（一本の染色体として一点交叉を行うこと）することは、記憶という直接的な表現型を遺伝子型と混合していることになる。

通常のダーウィンの進化論からすれば、進化は表現型に基づく自然選択によって起こる。個体の適合度を決定するのはその個体が現実に出した手であるから、「C」と「D」という選択肢が表現型であり、その記憶も表現型である。一方どういう手を出すかという指令は戦略領域で示されているのでその部分は遺伝子型である。そして表現型はそれをもたらす遺伝子型を通じてのみ次世代へ遺伝しなければならない。

したがって、ダーウィンの通常の進化論に基づくのであれば記憶領域と戦略領域は分別されねばならないし、戦略領域の64ビットだけが染色体として機能しなければならない。

Axelrod 実験における上述のような論点はこれまで論じられていない。結果として染色体における遺伝子パターンがある程度に収束しているため問題視されなかったか、¹⁵⁾ 平均適合度の観察だけで「協力の進化」に対する有効性を判断したためと思われる。

上述のように表現型と遺伝子型を混合して一つの染色体の中に入れてしまうことには原則的な混乱がある。したがって本稿ではダーウィンのコーディングを採用することにする。

15) 後述するようにシミュレーション結果から考えると、GAにおいてこのタイプの染色体が収束(ハミング距離がある程度以上に小さくなるという意味で)するとは思われない。50世代という特定の世代でGAを打ち切った理由はここにあると考えられる。おそらく染色体は収束しなかったのだ。

3.2 TFT パターン

Axelrod は実験結果の解釈に当たって「どのようなパターンを TFT とするか」という基準を示していない。彼が主張したのは2.2節で紹介した5つのパターンだけである。したがって最終的に得られた集団の何パーセントが TFT になったのかという量的な検討ができない。このことが GA 実験の後半部分で「変化する環境」によるシミュレーションを行った際に結論が曖昧になった原因になっている。

「1回前の相手の手を出す」という TFT の基本的な定義からすると、上述の5つのパターンのうち「DCDDCC に対して C」と「CDCCCC に対して C」は TFT とはいえない。前者は3回前に出された相手の C に対して2回前には D を出しているし2回前に出された相手の D に対して1回前には C を出している。後者は3回前に出された相手の D に対して2回前に C を出している。

5つのパターンすべてに共通するのは最後のゲームにおいて相手に出された手を次回にはまねをする、という点だけである（その意味ではすべて TFT に「似ている」）。

厳密に定義にしたがった TFT プレイヤーは得られた記憶領域が相手の前回の手をまねしたパターンになっていなければならない。すなわち

- ・記憶領域の第1ビット（3回前の相手の手）と第2ビット（2回前の自分の手）、第3ビット（2回前の相手の手）と第4ビット（1回前の自分の手）が同じ。
- ・「今度（4回目）に出す戦略（記憶領域のパターンで示される場所にある遺伝子）」が記憶領域の最終ビットと同じになっている。

Axelrod が見いだした5つのパターンは後者だけを共通としている。

記憶パターンとそれに対応した「次の手」だけで TFT を定義するのは、厳密に言えば不十分である。どういう戦略をもちいるか、というパターンは戦略領域のパターンで決まる。したがって、最終的に得られた集団の戦略領域のパターンがどのようなものを TFT と呼ぶのか基準をはっきりさせねばならない。個

体の染色体がその個体に TFT 戦略をとらせるようなパターンを明確に定義することは可能である。¹⁶⁾これを「遺伝的な TFT パターン」と呼ぶことができるが、後述のように遺伝的なパターンは検出できない。したがって以下では上にあげた「記憶パターンからみた TFT」の検出のみに焦点を当てる。

4 シミュレーション

以上で述べたように、Axelrod による GA 実験はコーディングにおける記憶の取り扱いと TFT の定義という2つの問題を抱えている。

そこで本稿ではあらたに染色体を定義し直し「ダーウィンの手法」によってシミュレーションを行う。またその際に TFT についての明確な定義を与えて結果を量的に分析する。

4.1 プログラムの概要

本稿で用いるシミュレーションプログラムのコーディングでは記憶を染色体に含めていない。

プログラム中で定義される個体は、戦略領域配列（長さ 64 の char 型）と記憶領域配列（長さ 6 の char 型）を持っており、個体が出す手は Axelrod のプログラムと同じく記憶領域配列を 2 進数と読み、それを 10 進変換して得られる数が示す個所の戦略領域配列から出される。

ゲームの度に記憶領域配列は更新される。各プレイヤーは集団中からランダムに選ばれた相手と対戦し、ゲームの繰り返し回数は平均 150、標準偏差 15 の正規分布から引き出される。囚人のジレンマゲームの利得構造は Axelrod が用

16) TFT 戦略をとる個体の記憶領域が示すパターンは「*xxyyz」であるから 16 通り (*, x, y, z はそれぞれ「C」「D」)。したがって戦略領域中でこれに対応する個所も 16 個所。具体的には「0, 6, 24, 30, 32, 38, 56, 62 ビットが C」「1, 7, 25, 31, 33, 39, 57, 63 ビットが D」となっていることが必要な戦略領域のパターンである。シミュレーションではこの「遺伝的な TFT」の検出も試みたが明確に検出することはできなかった。本文参照。

いたものと同じである。

繰り返しゲームが終わると、個体の1ゲーム当たり平均利得に基づいて「親」が選択され、交叉、突然変異というGA過程が行われるが、このとき操作対象となるのは戦略領域配列だけである。つまり染色体は戦略領域配列の長さ64の文字列のみからなる。染色体として戦略領域配列だけを操作対象とすることによって、遺伝子型と表現型の混乱は排除されている。

また、新しく生成された集団が最初のゲームを行う際には依拠すべき記憶がないので、各世代が最初に出す手を決めるためには仮想的な3回のゲームが必要である。ここではダーウィンのような進化をシミュレートしているので、この最初の記憶はランダムに生成されるものとしている。

4.2 TFTの定義

シミュレーションの結果得られる記憶パターンからその個体が「TFTかどうか」を判定するために以下のように定義する。

- ・記憶領域が「*xxyyz」の形をしており、
- ・それが指し示す場所の染色体の遺伝子座に「z」がある。

シミュレーションでは以上で定義したTFTパターンを検出する。

5 シミュレーション結果

以下のシミュレーションでは、Axelrod実験と同じく集団数20、交叉率0.25、突然変異率0.01を用いる。世代数の基準は500である。また、選択に際しては集団中でもっとも高い利得を得た個体は平均2回、平均的な利得を得た個体は1回、親として選択される¹⁷⁾交配は異性生殖とし、一点交叉によって2個の親から2個の子が作られるものとする。

17) 線形スケーリングを採用した。スケーリングアルゴリズムはGoldberg(1989)によった。

この実験を100回行い結果を検討する。

まず第1に「ラマルク的手法」と「ダーウィンの手法」の比較を行う。ラマルク的手法はダーウィンの手法に比較してTFTを発生させやすいことが明らかになる。

第2に集団の平均利得の推移を観察する。そこから「協力の進化」がダーウィンの進化によっても現れてくる可能性が高いこと、しかしそれは非常に不安定であり、協力の進化と崩壊を交互に繰り返すことを明らかにする。

最後にTFTの効果を考察する。TFTプレイヤーと対戦する確率をあげていくことによってTFT個体が集団中に占める比率は上昇する。すなわち、集団外部からTFT行動が与えられると集団は確実にそれを学ぶ。

5.1 Axelrod 実験との比較

Axelrodと同じくラマルク的手法でコーディングしたシミュレーションでは、遺伝的なTFTを示す個体は現れなかった。これについてはダーウィンの手法による場合でも同じなので¹⁸⁾以下では4.2で定義したTFTの出現率だけを問題にする。

100回の実験で最終500世代目の集団の中にいくつのTFT個体が現れたのかを両手法で比較したのがTable.2である。

階級値 (個体数)		0-5	6-10	11-15	16-20
ラマルク的手法	出現回数	49	15	14	22
ダーウィンの手法	出現回数	56	22	14	8
		平均	標準偏差		
ラマルク的手法		8.67	6.73561		
ダーウィンの手法		6.07	5.39121		

Table. 2 : TFT 個体の出現頻度

18) 実際には数十回のシミュレーションにおいて脚注16)で述べた遺伝的TFTを観測したが数が少なすぎる。

ラマルク的手法でコーディングした場合、100回の実験中で最終集団におけるTFTの出現頻度が16—20個体、つまり80%—100%になった回数は22回であるのに対して、ダーウィンの手法ではそれが8回しかない。平均値もあきらかにラマルク的手法の方が高い。

この結果からラマルク的手法を用いた場合にはTFTが集団に広まりやすいことがわかる。つまりAxelrodのコーディングでは、そもそもGAにおける染色体の設計手法がTFTになりやすいものであったのだ。記憶パターンからTFTを定義しているのだから、そのパターン自身が遺伝するラマルク的手法においてはそれは当然のことである。だからAxelrodの「TFTに似た戦略が進化した」という結論は染色体の設計自体に依存したものである可能性が捨てきれない。

この結果は、GAにおいて染色体の設計が非常に重要であることのよい証左を提供している。

5.2 協力の進化

TFTはダーウィンの手法において進化しにくいことを前節の結果は示唆している。それではダーウィンの進化においては「協力の進化」は起こらないのであろうか？

そうではない。

「協力の進化」という問題で重要なのはTFTという特定の戦略よりも「結果として協力が成立するように集団が進化するか」ということである。

例えばあるシミュレーションでは最終世代のTFT個体数が15であったがそのほとんどは「DDDDDDに対してD」というものであった。つまりTFTではあるが協力を進化させてはいない。

したがって集団の平均利得についての世代毎の推移から協力の進化を検討しよう。集団の平均利得が3に近ければ集団内で協力が普遍的になっていることを意味する。

結論からいえば世代数を十分に長く取るとダーウィンの進化は協力を進化させる可能性が高い。だが常にそうだとはいえない。またいったん集団内で協力が進化してもそれはひどく不安定で、すぐに崩壊してしまう。そしていったん崩壊してもふたたび協力へ向けて進化が始まる。

合計500回のシミュレーションで協力が進化した¹⁹⁾のは141回、約28.2%である。これからは「必ず協力が進化する」とはいえない。

この点からもAxelrod実験が不十分であったことがわかる。Axelrodは50世代のGAを10回行って結論を引き出しているが、10回という少ない回数では「協力が進化した」と結論づけるには不十分なのである。

他方、ラマルク的手法に基づく実験では合計500回のシミュレーションで協力が進化したのは401回、80.2%である。

この結論は明白である。ラマルク的手法は協力も進化させやすい。したがってAxelrodの結論はこの部分でも彼の取ったコーディング手法(ラマルク的手法)に大きく影響されている。

Fig. 1は協力が進化しなかったケースでの集団平均利得の推移をプロットしたものである。高い平均利得から出発したこの集団は300世代辺りまで平均利得を減少させている。320世代あたりで最低の1.0近くまで減少した後、370世代まで急速に平均値をあげていったがその後は再び緩やかに減少している。

一方Fig. 2は集団が協力へ進化したケースである。このケースでは最初に2.2付近から始まった平均値は200世代辺りまで跛行し、その後急速に上昇して267世代で集団は協力へ収束した。このときすべての個体が「CCCCCCに対してC」という形のTFT個体になっている。

協力へ収束したほとんどの場合に、集団中の個体は上記のような「CCCCCCに対してC」になる。このことは、同時になぜ「遺伝的なTFT」が進化できな

19) 「協力が進化した」ことの判断基準は、集団が協力へ収束した場合と、収束しなくとも最終世代の平均利得が2.9を越えていることである。なお集団の協力への収束は直近10世代のうち5世代以上で平均利得2.9を越えた時点で判断している。

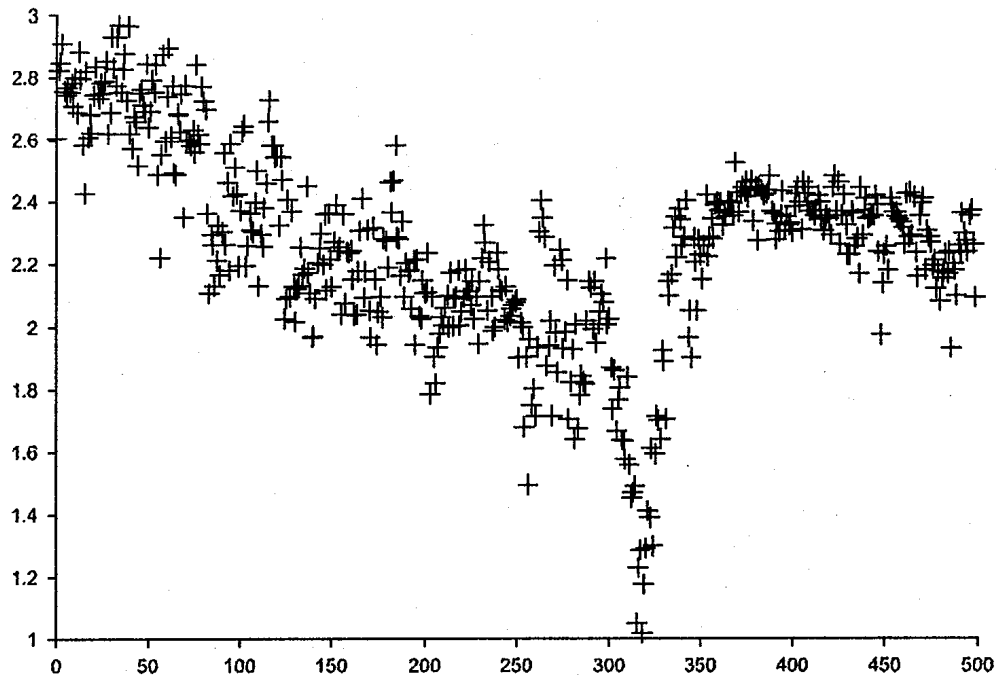


Fig. 1 : 協力が進化しないケース

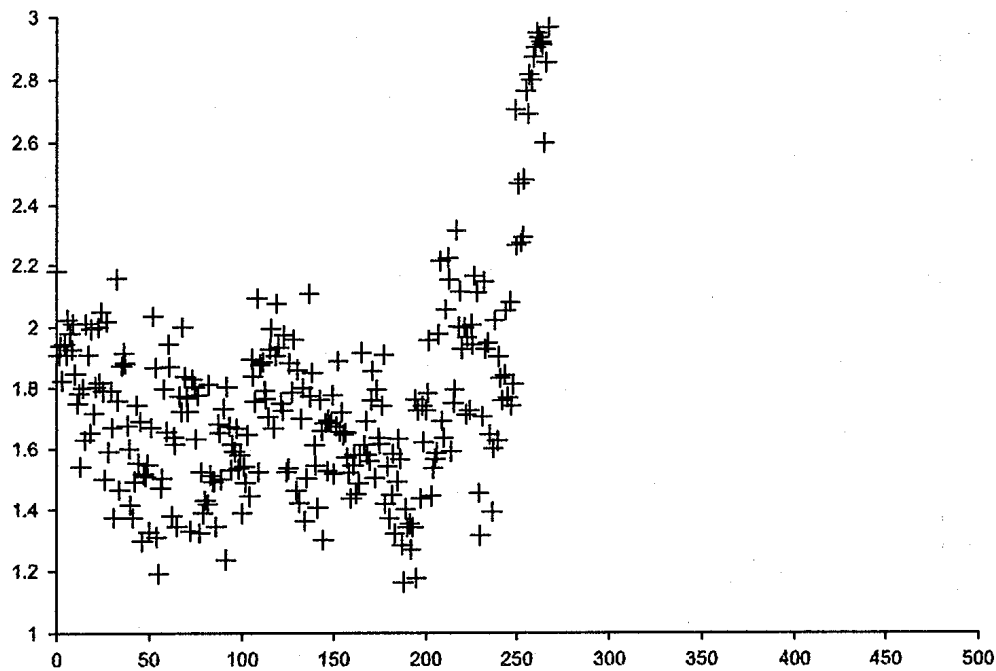


Fig. 2 : 協力へ収束したケース

いのかの理由も説明している。というのは記憶領域パターン「CCCCCC」に対して使われる遺伝子座は戦略領域配列の第0ビット目だけであり、残りの染色体は戦略の決定になんら関わりがない。したがって、遺伝的に TFT を定義した

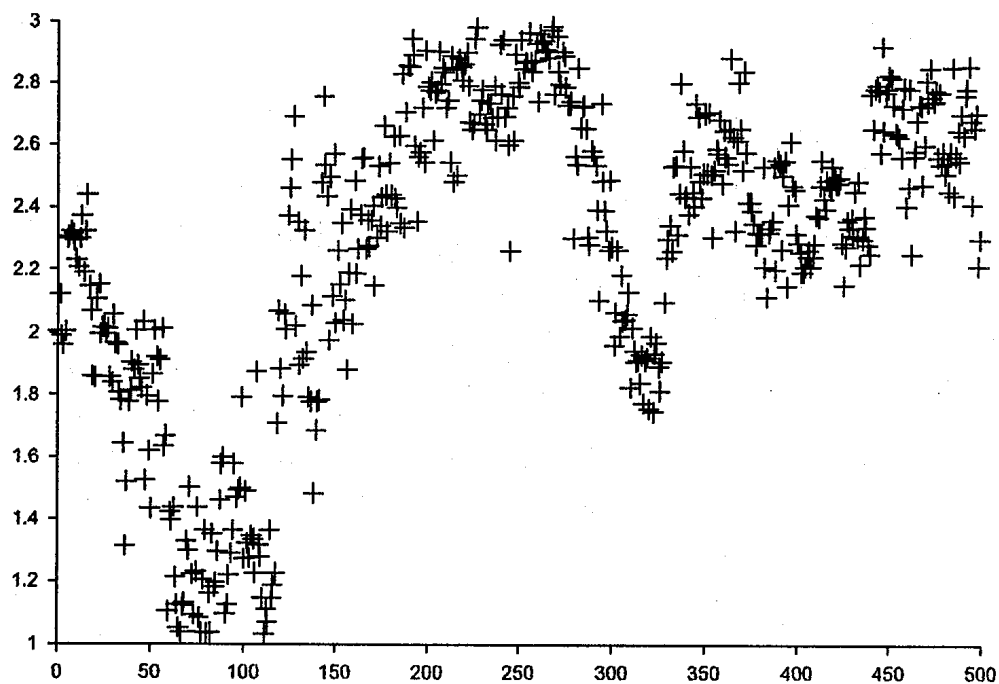


Fig. 3 : 協力の進化と崩壊

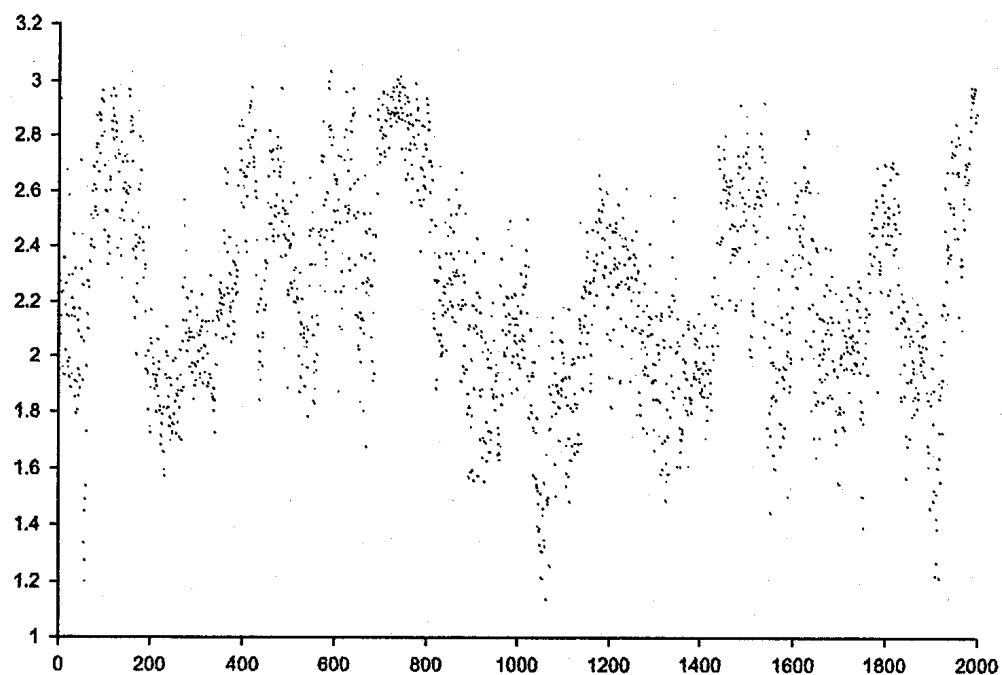


Fig. 4 : 長いスパンでの進化例

場合に TFT と見なすための特定ビット (脚注 16) 参照) は個体の行動に無関係となるからだ。さらに遺伝的な TFT は特定の遺伝子座に特定の文字がなければならないが、そのパターンは非常に長い。このため交叉によって容易に破

壊されてしまうのである。

いったん協力が進化しても、その協力が長く維持されることはほとんどない。Fig. 3 は協力が進化した世代を越えて GA をすすめた場合の典型例である。

集団の平均利得は 100 世代付近でいったん最低に落ち込んだ後上昇し 250 世代でほぼ協力へ進化している。しかしその後再び平均値は減少、320 世代あたりをボトムに再び上昇を始めるが、今度は 2.6—2.8 付近でとどまっている。

協力が崩壊しても、世代の推移にしたがって再び集団が協力へ進化することもある。Fig. 4 は世代を 2,000 世代まで延長した典型例である。この集団は 100, 400, 500 世代あたりで協力を進化させ、2,000 世代で再び協力を進化させた。この例にあるように集団の進化を長いスパンでながめると、平均利得はこのような循環を繰り返す。集団が協力に必ず進化するかどうかは厳密にはいえないにしても、この循環の中で協力が進化してくる可能性が非常に高い。

平均利得が循環的な傾向を示す理由は以下のようなものである。まず集団が協力を進化させたとする。その中に「裏切り」戦略をとる個体が現れたらその個体は高い利得を得るのでその遺伝子が集団内に拡大する。すると集団の平均利得は低下していく。協力に対して裏切る個体が増加するほど平均利得は低下していく。この過程がそのまますすめば Fig. 1 の 320 世代付近で見られるように集団は「裏切り戦略」へ進化していくが、ここでも突然変異によって互恵的な個体(相手が協力したら自分も協力する個体)が現れて彼ら同士の対戦によって高い利得が獲得されれば再び集団は協力へ向かう。

この過程において特徴的なことは、協力へ向かうにせよ裏切りへ向かうにせよ、一時期集団は同じ方向へ進化する、つまり定向進化を起こる傾向があることである。これは突然変異が比較的起きにくいことに原因を求めうる。有効な突然変異が起きなければ集団の進化方向は変化しない。比較的短いスパンでこのような突然変異が起こった場合は Fig. 3 の 320 世代付近で見られるように裏切り戦略へ収束することなく進化方向は反転する。

遺伝子パターンが集団内で収束せずに循環的な進化を起こす点は、GA が関

数最適化で使われる場合と大きく異なる点である。最適化問題の解法として GA が使用される場合、集団の遺伝子型はある一定のパターンに収束する。その収束によって我々は最適解へ到達したことを知るのである。しかしここで行った GA のように、利得を規定する表現型があるかぎられた遺伝子座で決定されている場合には集団の遺伝子パターンは収束することなく進化過程を繰り返す。

5.3 TFT の影響力

以上のシミュレーションはいずれも環境を固定しない場合の実験である。ここでは各個体は集団中の他の個体とランダム対戦を繰り返す。この場合対戦相手は自分と同じく記憶パターンから「手」を決定しているのでなにかの戦略に特化していると言うわけではない。

個体が対戦する相手がある戦略に特化している場合は環境が固定されていることを意味する。上述したように Axelrod の GA 実験の大部分は環境を固定して行われている。

以下では対戦相手の戦略が TFT に固定されている場合を取り上げる。これは集団の外部から TFT という戦略が与えられていることを意味する。ダーウィンの進化を行う集団は外部環境として TFT が与えられると協力を進化させやすくなり、また集団内部で TFT を進化させやすくなる。

以下のシミュレーションでは、個体が TFT 戦略者と対戦する確率を 0% から 100% まで 5% 刻みに変化させ (以下ではこれを「TFT 濃度」と呼ぶ)、結果を比較する。この確率が 0% ならば集団は全くの可変環境にあり、100% ならば完全に固定された環境である。

まず、協力の進化についてみてみよう。

集団が協力を収束した場合²⁰⁾はそこで GA を打ち切り、収束しなかった場合

20) 収束の定義は脚注 19) 参照。

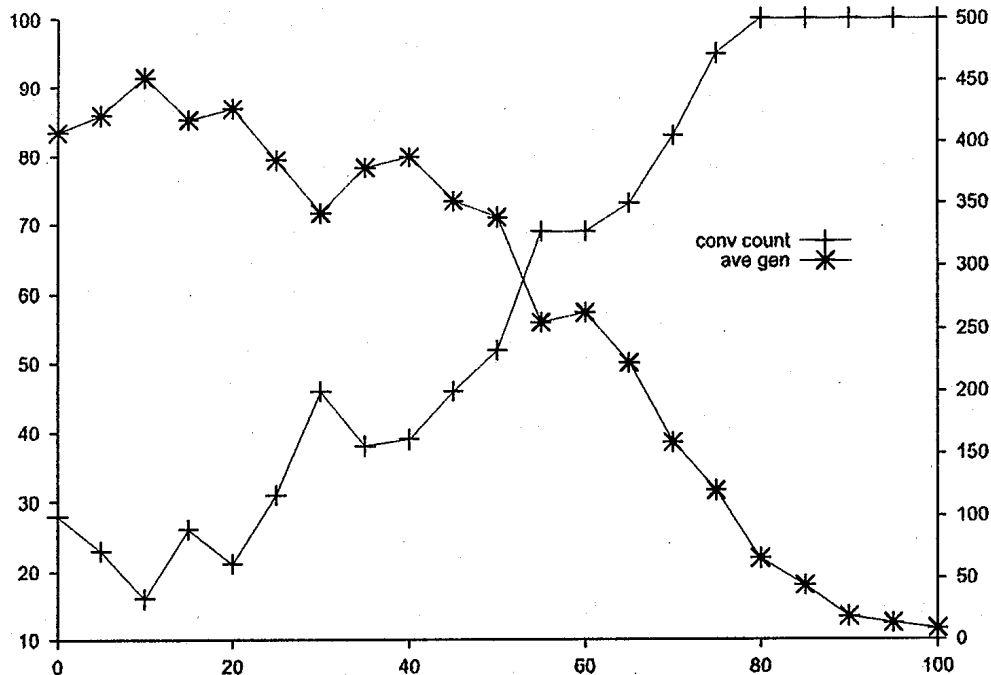


Fig. 5 : TFT 濃度と協力の進化

は500世代までGAを続けることにして、100回の実験中収束した回数と収束平均世代²¹⁾をTFT濃度毎にプロットしたのがFig.5である²²⁾

TFT濃度が80%を越えたシミュレーションでは100回の実験すべてが協力を進化させた(グラフ左目盛り)。濃度の低下に伴って協力が進化してくる回数は減少し、TFT濃度が0-10%では100回中15-20回ほどしか協力は進化してこなかった。

また、平均的な収束世代もTFT濃度によって変化する(グラフ右目盛り)。濃度が高い場合は比較的早い段階で集団は協力へ収束する。例えばTFT濃度100%、すなわち個体がTFT戦略者としてしか対戦しない固定環境の場合には平均収束世代は8.47世代であり、きわめて急速に協力を進化させる。それに対して濃度が0-20%では平均収束世代は400世代を越えた。

以上の結果からTFT環境が協力の進化に与える影響は明白である。ある程

21) 収束しなかった場合は最終世代を収束世代と見なす。

22) 図中「conv count」とあるのが100回の実験中で収束した回数であり左目盛り。「ave gen」とあるのが平均収束回数であり右目盛り。

度の量の TFT が外部から与えられると集団は協力を進化させやすくなる。したがって、ダーウィンの、過去の記憶を遺伝していない集団でも外部環境から TFT 戦略が与えられればそれを模倣する傾向がある。

このことは TFT 戦略が遺伝的な意味で侵入しやすい、ということの意味するものではない。このシミュレーションでは各 GA の中で TFT 濃度は固定されており、TFT 濃度自体が進化過程の中で増加するわけではないからだ。ここでの結果はむしろ GA 過程が外部環境を模倣する、あるいは学習する性質を持っていることを示している。

次に TFT の進化をみてみよう。

協力が進化した時点で GA を止めると当然その時には集団は TFT 個体で収束している。したがって集団中に発生した TFT の数を調べるには、一定の時点での集団構成を比較する必要がある。ここでは集団の進化を 500 世代で停止させた時点での TFT 個体が集団中に占める比率を調べる。

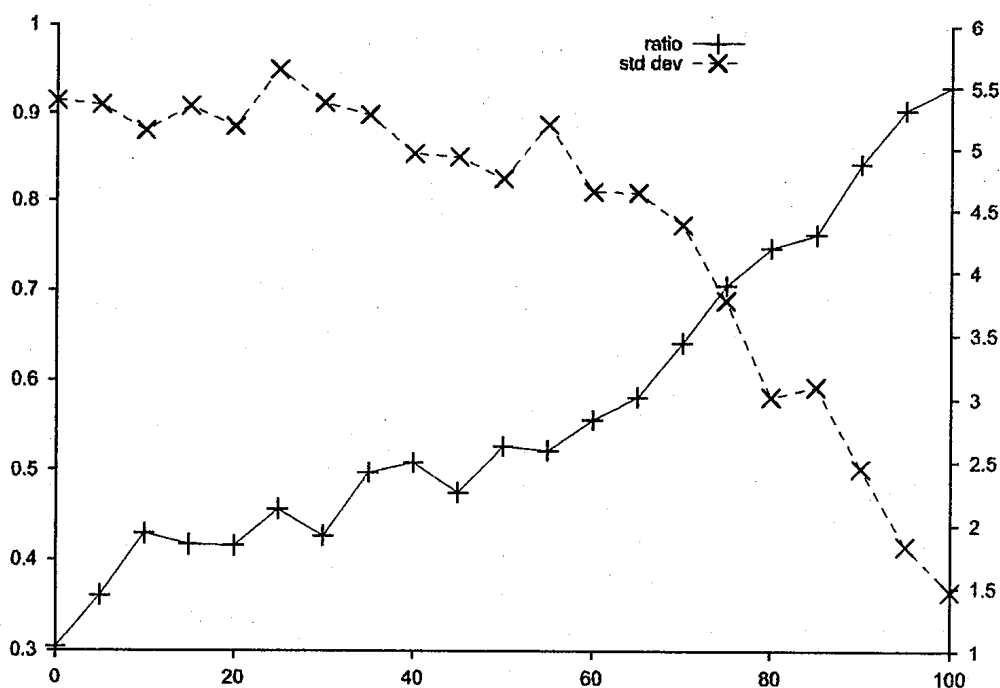


Fig. 6 : TFT 濃度と集団中の TFT

23) 図中「ratio」が割合で左目盛り、「std dev」が標準偏差で右目盛り。

各 TFT 濃度で 500 世代 GA を 100 回繰り返す。最終世代での TFT 個体が集団中に占める割合と、個体数についての標準偏差を示したのが Fig. 6 である。²³⁾

あきらかなように、TFT の平均割合 (左側目盛り) は TFT 濃度が上がるにつれて上昇し、他方 TFT の標準偏差 (右側目盛り。個体数の標準偏差であることに注意) は減少している。このことは TFT 環境が集団を TFT が進化するように導くことを示している。

以上のように、TFT 環境が与えられ、それとの対戦確率が増えるにつれて、協力も TFT もどちらも進化しやすくなることがわかる。

6 結 論

Axelrod が GA 実験において TFT の進化を結論づけた際には、それが固定環境下での実験であったこと、TFT の定義が曖昧であったこと、コーディング手法自体が TFT を進化させやすいものであったことによりかなりな程度依存していたように思われる。

本稿のシミュレーションでみたようにダーウィンの進化を想定した場合も、十分に長いスパンでは協力は進化するし、したがって TFT も発生し集団中に普及する。

しかし、ここで検出する「TFT 戦略」はあくまでも事後的な記憶パターンから見たものだけであることに注意しておきたい。TFT は戦略であって必ずしも協力を意味するものではない。したがって、協力が進化すれば記憶パターンは TFT を示すが、逆に集団の多くの個体が TFT だから協力が進化した、とはいえない。

しばしば言及したように、ある個体に TFT 戦略を取らせる遺伝子型は定義できるのであるが GA の中では現れにくい。ここで用いた染色体のコーディングでは TFT 戦略を個体にとらせる遺伝子パターンは非常に長い。Holland

(1992)の「積み木理論」によると、短くて効果が高い遺伝子(染色体の一部)が進化の過程で蓄積され、それが有効な表現型をもたらす遺伝子型を進化させる。それに照らすと長いパターンを必要とするような遺伝子型は進化しにくいのである。

この点はGAにおける「戦略」決定の方式に深く関係している。例えばラマルク的手法であれ、ダーウィンの手法であれ、直前の相手の手(記憶領域の第5ビット)が「C」であれば、それに対応する可能性がある染色体は「偶数ビット」ということになる。そして偶数ビット、というように長い染色体は次の交叉で必ず破壊されてしまうことになる。したがってGA実験からはTFTを支える遺伝子パターンの検出は難しい。

遺伝的なTFT戦略自体の進化を研究するためにはここで用いたように「記憶配列を2進数と読んでそれが指し示す戦略領域の場所」というような単純なものではなくもっと巧みな手法が必要とされよう。それは別の課題となる。

ダーウィンの手法で行われるGAのように、非常に単純な過程においても、外部からTFT環境が与えられると、それに応じて協力を進化させるという結論はこれまで指摘されてこなかった点である。

RPDとTFTに関する研究は「進化的に安定な戦略(ESS)」やフォーク定理との関連へ関心が向けられてきた印象がある。この場合、ESSの概念は非常に狭く、フォーク定理から導かれる均衡概念は非常に広い²⁴⁾ので、「中間の」均衡概念を構築する方へ研究がすすんだのである。

したがって本稿でとりあげたような単純な進化機構をもった集団に対してTFTという環境が持つ影響力というような問題については誰も関心を持たなかったものと思われる。

このシミュレーションで示された「協力の進化」に関する結果は重要である。

24) ESS基準からするとRPDにはそのような戦略はなく、フォーク定理からすると協力を含んで無限の均衡解が出現する。

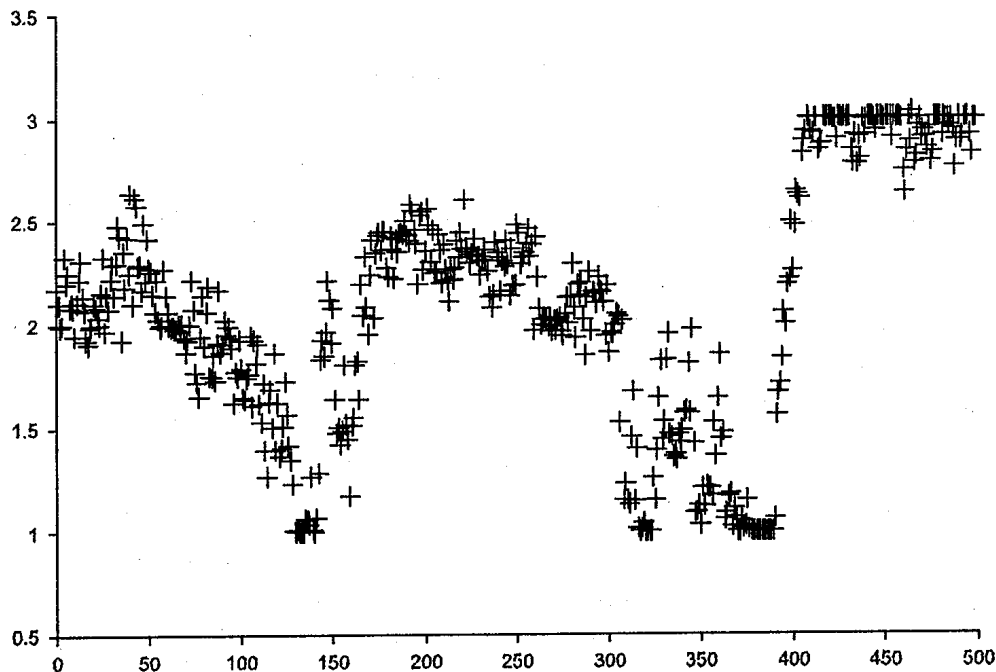


Fig. 7 : ラマルクの進化の一例

ラマルク的手法とダーウィンの手法での比較からは「協力の進化」において「記憶の遺伝」が大きな影響を持っていることがわかる。また、今回は十分な検討ができなかったため断定はできないが、平均利得の変化グラフをみると、ラマルクの進化の場合、協力はいったんそれが集団中に進化すれば（ダーウィンの進化に比較して）粘着性を持つような印象を受ける。典型例を Fig. 7 にあげておく。この例では 350 世代辺りで 40 世代にわたって「裏切り」が持続した後急速に協力を進化させ、100 世代近くおなじような集団を維持している。このような粘着性はダーウィンの進化ではみられなかったものである。社会進化はラマルク的な進化に支えられると考えられる。したがって記憶を次世代に引き継ぐことが、RPD のように個人利益と社会利益が対立するようなシチュエーションにあっては非常に重要なファクターとなる。

やや曖昧な言い方になるが、社会がコンフリクトを少なくし、協力を進化させるためには過去の歴史を次世代へ引き継ぐことの重要性をささやかながら示唆しているともいえよう。

References

- 伊庭齊志 (1994), 遺伝的アルゴリズムの基礎—GAの謎を解く—, オーム社。
- Axelrod, R. (1980a), 'Effective choice in the prisoner's dilemma', *Journal of Conflict Resolution* 24 (1), 3-25.
- Axelrod, R. (1980b), 'More effective choice in the prisoner's dilemma', *Journal of Conflict Resolution* 24 (3), 379-403.
- Axelrod, R. (1984), *The Evolution of Cooperation*, Basic Books, New York.
- Axelrod, R. (1987), The evolution of strategies in the iterated prisoner's dilemma, in L. Dabiz, ed., 'Genetic Algorithms and Simulated Annealing', Pitman, pp. 32-41.
- Boyd, R. & Lorberbaum, J. P. (1987), 'No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game', *Nature* 327, 58-59.
- Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.
- Hoffmann, R. (2000), 'Twenty years on: The evolution of cooperation revisited', *Journal of Artificial Societies and Social Simulation* 3 (2), <<http://jasss.soc.surrey.ac.uk/JASSS.html>> .
- Holland, J. H. (1992), *Adaptation in Natural and Artificial Systems*, 2nd edn, MIT Press.
- Linster, B. G. (1992), 'Evolutionary stability in the infinitely repeated prisoners' dilemma played by two-state moore machines', *Southern Economic Journal* 53, 880-903.
- Michalewicz, Z. (1990), *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd rev. and extended edn, Springer.
- Mitchell, M. (1996), *AN INTRODUCTION TO GENETIC ALGORITHMS*, The MIT Press, Cambridge.
- Nowak, M. A. & Sigmund, K. (1992), 'Tit for tat in heterogeneous populations', *Nature* 355, 250-253.