Vol.2012-MPS-88 No.3
2012/5/17

進化型繰り返し囚人のジレンマにおける最適戦略の探究

糸井 良太[1,a)]　田中 美栄子[1,b)]

概要：戦略の自動進化を取り入れた繰り返し囚人のジレンマモデルにおいて，特に 2 進表現による 1 次元遺伝子配列の 2 倍変異や 2 分裂変異を点変異と組み合わせることで遺伝子の長大化の効果を考察する事を目的とした Lindgren モデルがある．我々はこのモデルに依拠して長大な遺伝子配列が出現するまでシミュレーションを行うことにより，長期間生存する遺伝子配列の内部構造が，しっぺ返し戦略とパブロフ戦略，およびその亜種などの特定の遺伝子配列を要素として持つことを見出した．これらの戦略は単純なしっぺ返し戦略より強く，多様な戦略との対戦に勝利した結果，長期間生存し続けることのできるロバストな戦略であると言える．このような戦略の出現・生存に対する条件について考察する．

キーワード：囚人のジレンマ，戦略の進化，進化計算

A Study on the Optimum Strategy in the Iterated Prisoner's Dilemma with Evolvable Genes

Ryota Itoi[1,a)]　Mieko Tanaka-Yamawaki[1,b)]

Abstract: In the realm of iterated prisoners' dilemma equipped with evolutional generation of strategies, a model has been proposed by Lindgren that allows elongation of genes by means of doubling and fission of one-dimensional genetic arrays multiplied with point mutation of genes. We conducted numerical simulations of this type of models under various conditions, and found that long-lived strategies with long genetic array have some particular elements such as Tit-For-Tat (TFT), Pavlov (PVL), and Ritaliation-oriented-Tit-For-Tat (R-TFT) in common. Such strategies are stronger than a simple TFT, and robust strategies that survive under many matches with various kinds of strategies. We consider when and how such strategies are generated in the history of simulation.

Keywords: Prisoner's Dilemma, Evolution of strategy, Evolutional computation

1. はじめに

ゲーム理論は，互いに影響を与え合う複数の主体の間で生じる戦略的な相互関係を研究するためのツールであり，人間行動や経済の動きを理解するための手法として重要である．2 人ゲームの最も良く知られた例である「囚人のジレンマ」は，2 人のプレイヤーの合理的選択が，全体では良い結果にならない様子をモデル化したもので，二酸化炭素排出問題や核保有問題，また価格競争など，身近な問題を考える際にもモデルとして役立つものであり，昔から研究されてきた [1][2][3]．歴史的に有名なアクセルロッドの実験においてはしっぺ返し戦略が最強であったが [4]，環境を変えると最強戦略もまた変化することや，また最初からしっぺ返しが存在したのか，何らかのプロセスを経て発生してきたのか，また何らかの要因により消滅するのか等の問題を考えると，もっと視点の時空を広げ，人工生命的な観点で様々の戦略の生成消滅する様子を観察することは大変興味のある研究課題と言える．

[1] 鳥取大学大学院工学研究科エレクトロニクス専攻
Tottori University, Graduate School of Engineering,Department of Information and Electronics
[a)] s072009@ike.tottori-u.ac.jp
[b)] mieko@ike.tottori-u.ac.jp

戦略を 0,1 列からなる遺伝子と見立て，遺伝的アルゴリズム（GA）等を用いて自動進化させることによって，強い戦略を自動生成させようという試みは，Lindgren[5] によって導入された遺伝子列の 2 倍化と 2 分裂を点変異と組み合わせることで，より環境に適応した長い遺伝子列を創成しようとする，人工生命的試みへと発展していった．

一方，村上等 [6] は，しっぺ返し戦略を最初から存在したものではなく，適者生存のルールによって自動生成されるというモデルを考察し，シミュレーションによってしっぺ返しが生成され生存する様子を観察した．

我々は Lindgren モデルの初期条件を変えて長期間シミュレーションする途中，カンブリア紀になぞらえることのできる，多数の異なる遺伝子が発生する時期のあることを見出し，その生成消滅の様子と，エントロピーの増減とを結び付けて，このような事象の発生する条件について考察した [7].

本稿では再び Lindgren モデルのシミュレーションの結果から，生成された長い遺伝子列の持つ特徴を調べることで，生存する長い遺伝子には共通した特徴があり，実はその殆どが，0001（報復力の強いしっぺ返し戦略），1001（通常パブロフ戦略とよばれるもの），そして 01=0101（しっぺ返し戦略）を要素として持ち，主としてこれらの組み合わせからなる遺伝子であることを見出した結果を報告する．

2. 囚人のジレンマ

2.1 基本概念

囚人のジレンマは一般的に表 1 の利得表が用いられる．パラメータは R，P，S，T の 4 つで，$S+T<2R, S<P<T<R$ となるように設定されている．これの意味するところは，

- 相手が協力を行うと仮定したとき，自分の得られる得点は自分が協力した場合に得点 R を得る一方，裏切れば得点 T を得る．T の方が R より大きいので裏切り行動を選択するのが合理的である
- 相手が裏切りを行うと仮定すると，自分の得られる得点は自分が協力した場合に得点 S を得る一方，裏切れば得点 P を得る．P の方が S より大きいので裏切り行動を選択するのが合理的である

つまり，相手の行動に関わらず自分の最も合理的な選択は裏切り行動になる．当然，相手も合理的な選択を行うならば裏切り行動になり，この場合の両者の利得はいずれも P となってしまい，両方で協力し合った時の利得 R より小さくなってしまう．損を承知で裏切りあうほかない，というのがジレンマなのである．

表 1 囚人のジレンマの利得表
Table 1 Payoff table of Prisoner's Dillemma

自分，相手	協力	裏切
協力	R , R	S , T
裏切	T , S	P , P

※ S+T<2R,S<P<T<R

2.2 繰り返し囚人のジレンマ

1 回だけの囚人のジレンマにおいてプレイヤーはお互いに裏切りあってしまう．しかし，終わりを告げずに囚人のジレンマを繰り返し行う繰り返し囚人のジレンマ (Iterated Prisoners Dilemma：以下 IPD) を行うことによって．協力行動を行う戦略でも勝ち残る事がアクセルロッドの実験によって証明された．また，優秀な戦略には以下に示す 3 つの性質を持っていることが確認された．

- 自分からは裏切らない
- 相手の裏切りにはすぐに裏切りで反撃する
- 相手が協力してくればこちらもすぐに協力する

また，最も成績の良い戦略であった「しっぺ返し戦略」は，この 3 つの性質を兼ね備えていた [4].

3. 進化型 IPD モデルにおける戦略の自動進化

3.1 進化型 IPD モデル

Lindgren はマルチエージェントモデルに於いて，戦略を生物の遺伝子に見立てて 1 次元のバイナリ文字列で表し，それらが利得を評価値として遺伝的アルゴリズムに従って進化する状況をモデル化した [5]．これは人工生命研究の中で注目され，多くの研究者の興味を引き付けた [8][9][10][11]. Lindgren モデルは，エージェントに行動決定の指針である戦略を持たせ，その戦略を進化，退化，淘汰をいった要素を用いることで自動進化させていくモデルである．このモデルを用いることで，生物が単純な生物から現在のように非常に多様で複雑に進化してきたように，戦略を進化させることができ，複数の戦略を同時に使用してシミュレーションを行う場合，戦略の組み合わせは無限にあるが，このモデルを使用ことで戦略の検索領域を絞ってシミュレーションを行うことができる．本研究では，このモデルを用いて実験を行った．

3.2 戦略の 2 値表現とノイズ

戦略は裏切りを表す’0’と協力を表す’1’の 2 値文字列で構成され，高さ m の 2 分木の葉として表現される．m は参照する歴史（過去の自分と相手の行動）の数であり．m=1 は直前の相手の手のみを考慮する戦略となり，高さ 1 の 2 分木となる．この場合の可能な戦略は，00，01，10，11 の 4 種類のみで，00 は相手の行動見よらず何時も裏切り，逆に 11 は何時も協力する手である．01 は直前の相手を同じ手を模倣する戦略で，しっぺ返し戦略（TIT for TAT:TFT) である．10 はその反対の手を出すので逆しっぺ返し（Anti Tit for Tat:A-TFT）である．m=2 は直前の自分の手までを考慮する場合に相当する．図 1 に 1101 戦

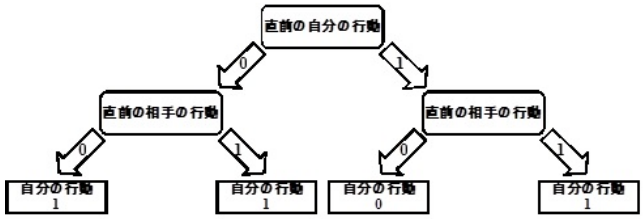

図 1　過去の手と戦略の関係 (m=2 の例)
Fig. 1　A strategy and the history of actions (m=2)

略を例として示した．ただし，今回の実験では m の最大値は 5 としている．さらに，必ずしも戦略通りに行動するわけでなく，一定確率で行動選択を失敗する．この失敗をノイズと定義する．

3.3　人口動態

戦略は人口という属性を持ち，自身を含めた全ての戦略と総当たりで対戦を行い, 獲得した利得の平均値と戦略全体での利得の平均値の差を評価値として，人口を変動させていく．文献 [5] では，対戦回数が定数であったため，t 世代目の戦略 i の獲得平均利得 S_i は，戦略 i の人口比率を X_i (戦略 i の人口/全体人口)，戦略 i が戦略 j より獲得した利得を $g_{i,j}$ と定義すると，平均値の算出には以下の式を用いている．

$$S_i(t) = \sum_j^N g_{i,j}(t)x_i(t) \tag{1}$$

しかし，本研究で用いるモデルは，対戦終了を確率的に決定しているため，対戦回数が毎回異なる．よって，戦略略 i が戦略 j と対戦した回数を $c_{i,j}$ と定義すると，平均値の算出には以下の式を用いる．

$$S_i(t) = \sum_j^N \frac{g_{i,j}(t)x_i(t)}{c_{i,j}(t)} \tag{2}$$

また，戦略 i の人口比率 X_i (戦略 i の人口/全体人口)，戦略 i の獲得平均利得を S_i，全体の平均利得を $\bar{S}$ とした時の人口の変化は以下の式で表される．

$$X_i(t+1) - X_i(t) = \alpha(S_i(t) - \bar{S}(t))X_i \tag{3}$$

α は増加率を表す．式 3 より，$\bar{S}$ 以上の利得を得ている戦略は人口を増やし，$\bar{S}$ 未満の戦略は人口を減らすのが分かる．つまり，人口が戦略の強さの指標となる．

3.4　突然変異

次の 3 つの変異を一定確率で用いて，新しい戦略を出現させる．

- 点変異・・・戦略の 1 個所が'0'→'1' もしくは，'1'→'0' に反転する
- 複写変異・・・戦略情報の長さが 2 倍になる
- 分離変異・・・戦略情報の長さが半分になる

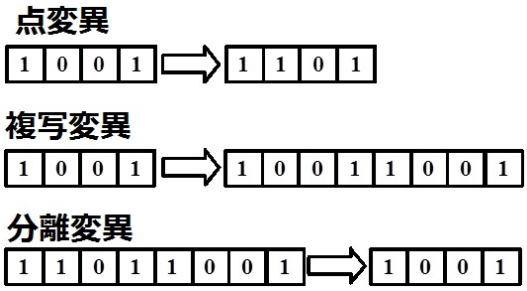

図 2　本稿で考察する 3 種の突然変異 (点変異，複写変異，分離変異)
Fig. 2　Three types of mutations used in our model (point mutation, doubling, separation)

表 2　実験条件
Table 2　Experimental condition

総人口	1000	増減率	10
対戦終了確率	0.5%	ノイズ発生確率	1%
点変異	2×10^{-3}%	複写変異	1×10^{-4}%
分離変異	1×10^{-4}%		

表 3　初期戦略
Table 3　Initial strategy

00	全て裏切り行動
01	直前の相手の行動と同じ行動
10	直前の相手の行動とは逆の行動
11	全て協力行動

表 4　利得表
Table 4　Payoff Table

自分，相手	協力	裏切り
協力	3 , 3	5 , 0
裏切り	0 , 5	1 , 1

具体的な例を図 2 に示す．複写変異時に追加される情報は元の戦略情報のコピーであり，分離変異は戦略を 2 分割し，どちらか一方をランダムに選んで新しい戦略としている．

4.　実験

4.1　実験目的と実験条件

第 3 章で述べた進化型 IPD モデルを用いて，戦略，人口分布が時間とともに変化する動的な環境における最適戦略の分析を行う．また，シミュレーションを行う上での実験条件，初期戦略，用いた利得表を表 2，表 3，表 4 にまとめる．利得表より，最大平均値はお互いに協力したときの 3 であり，最少平均値はお互いに裏切りあった時の 1 である．したがって，平均利得が 3 に近いほど協力関係が構築されているといえる．

4.2　動的環境における囚人のジレンマ

シミュレーションの結果，3 つの戦略が規則正しいパター

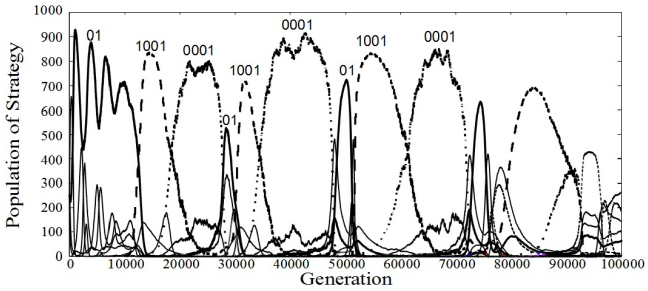

図 3　3 竦みの繰り返しパターン
Fig. 3　The triplet pattern (1001-0001-01) repeats in the long run

ンで出現する結果が得られた．動的な環境において，その他の戦略を差し置いてこの 3 つの戦略が何度も人口を大きく伸ばしていることから，この 3 つの戦略が有用な戦略ではないかと考えられる．そのパターンを下記と図 3 に示す．図は縦軸が人口数を表し，横軸が経過世代数を表している．

(1) 表 3 に示した初期戦略から 01 戦略（しっぺ返し戦略）が人口を大きく伸ばす．
(2) 01 戦略の代わりに 1001 戦略 (Pavlov 戦略) が台頭する
(3) 1001 戦略の代わりに 0001 戦略が台頭する
(4) 0001 戦略の代わりに 01 戦略が台頭する
(5) (2) から (4) を繰り返す

表 3 に示す初期戦略の中では，01 戦略が安定して勝つことが出来るため，01 戦略が人口を伸ばす．01 戦略が人口割合の大半を占めている時に，1001 戦略が出現すると，表 5 に示すように，自身同士の対戦の時，01 戦略はノイズによって C（協力）と D（裏切り）を交互に出すようになり，平均利得が 2.5 となるのに対して，1001 戦略はノイズが発生してもすぐに協力を出し合う関係に戻ることが出来る．そのため，平均利得は 3 のままである．また，1001 戦略と 01 戦略が対戦を行いノイズが発生しても得られる利得は共に同じである．つまり，1001 戦略は 01 戦略に勝つことが出来る．そのため，01 戦略の次に 1001 戦略が人口を大きく伸ばしている．しかし，1001 戦略は，表 5 に示したように，相手が裏切ったとしても協力を行ってしまう．そのため，0001 戦略のように裏切り行動を行い易い戦略に非常に弱い特徴を持つ．そのため，0001 戦略が出現すると同時に衰退し始める．0001 戦略は自分か相手が裏切りなら，必ず次手が裏切りになる報復思考の強い TFT 戦略 (Ritaliation oriented Tit For Tat:R-TFT) と言える．そのため，ノイズのある環境では，協力行動を続けることが難しい．逆に，01 戦略は裏切りやすい戦略に対して，得点を取られにくい特徴を持つため，0001 戦略の次に 01 戦略が台頭したと考えられる．

4.3　戦略の遺伝情報についての検証

図 3 に示すように同じパターンを繰り返すだけでなく，

表 5　ノイズが 1 回発生した時の 01 戦略と 1001 戦略の行動の変化
Table 5　Behavioral changes of 01 strategy and 1001 strategy when the noise occurs once

TFT	VS	TFT	1001	VS	1001
C		C	C		C
C		'D'	C		'D'
D		C	D		D
C		D	C		C
D		C	C		C

※' 'はノイズを表す

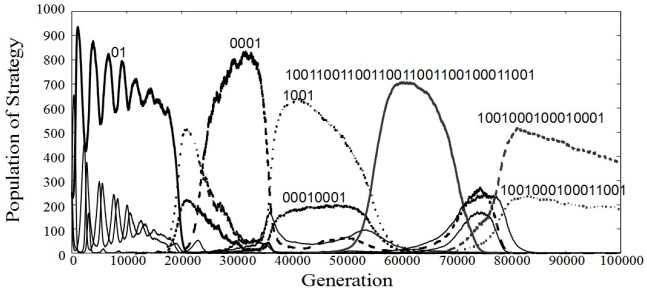

図 4　長く生き残る戦略は 1001 戦略と 0001 戦略を要素として含む
Fig. 4　Surviving strategies tend to contain 1001 and 0001 as their ingredients

1001 戦略・0001 戦略の 2 つの戦略は長く生き残る戦略の遺伝子情報として頻繁に現れる．その例を図 4 に示す．00011001 戦略や 1001000100011001 戦略など，遺伝情報に 1001 戦略や 0001 戦略を含んでいる戦略の人口が増大している．このことから，1001 戦略・0001 戦略を遺伝子に持つ戦略は，動的環境において長期的に勝ち残ることが出来るのではないかと考えられる．

長期的に勝ち残る戦略が，1001 戦略や 0001 戦略を遺伝子に含んでいるかの検証を行った．戦略長 4 以上の戦略を対象に，各戦略を長さ 4 の遺伝子情報で区切り，その遺伝情報が 0000 から 1111 のどの遺伝子情報を受け継いでいるかを生存期間別も調査を行った．その結果を図 5 から図 8 に示す．各図の右軸は生存していた期間を表し，縦軸はその期間に各遺伝子情報を持つ戦略がいくつ存在していたかの数を表す．

図 5 に示すように，0101 戦略は 5 万世代以上生き残る戦略の遺伝子にしっかりと出現している．しかし，図 6 と図 7 を見ると，図 5 と以上に，5 万世代に 1001 戦略や 0001 戦略を遺伝情報に持つ戦略が現れていることが確認できる．このことから，1001 戦略と 0001 戦略を遺伝情報に持つ戦略は 0101 戦略を遺伝情報に持つ戦略よりも多いことが分かった．図 8 より，その他の戦略は 5 万世代以上生存した戦略の遺伝子情報にはほとんど出現していないことが分かった．しかし，0000 戦略を遺伝情報に持つ 5 万世代以上生き残る戦略は多数存在していることが確認できる．このことから，5 万世代以上生き残る戦略を構成する遺伝子は，0101・0001・1001・0000 の 4 つであることが分かる．

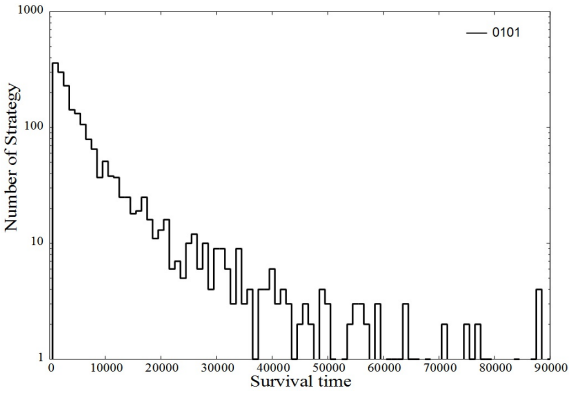

図 5　TFT 戦略は長期間生存する戦略の遺伝子に含まれるが，pavlov や R-TFT に比べると少ない

Fig. 5　TFT strategy is contained in the long lived strategies. but, few than Pavlov or R-TFT

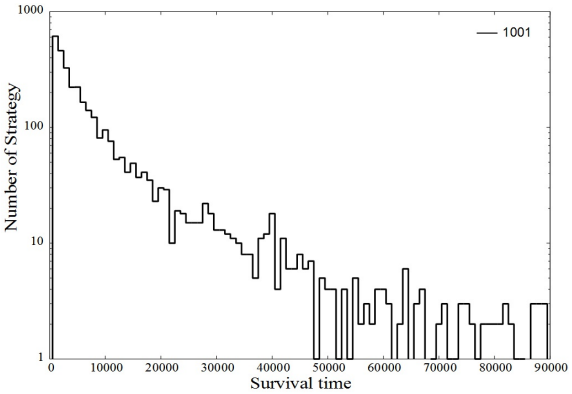

図 6　Pavlov 戦略は長期間生存する戦略の遺伝子に多数含まれる

Fig. 6　Pavlov strategy is contained many in the long lived strategies

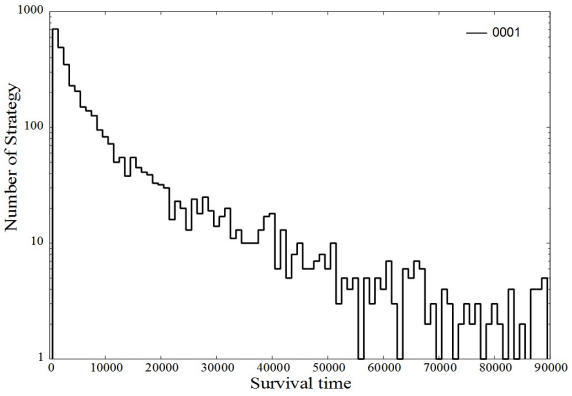

図 7　R-TFT 戦略は長期間生存する戦略の遺伝子に多数含まれる

Fig. 7　R-TFT strategy is contained many in the long lived strategies

また，表 6 に示す長期間生存していた戦略の例からも，長期生存戦略の遺伝子情報が 0101・0001・1001・0000 の 4 つのみであることが確認できる

4.4　実験結果及び考察

動的環境での囚人のジレンマの結果，図 3 に示すように，

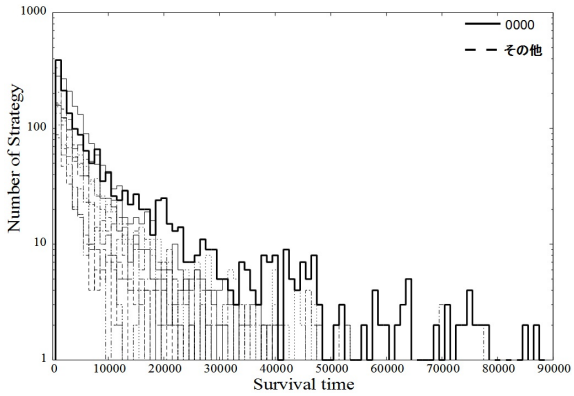

図 8　ALLD 戦略は長期間生存戦略に含まれるが，その他の戦略は 5 万世代を境に含まれない

Fig. 8　ALLD strategy contained long live strategies of gene. but, other other strategies not contain

表 6　長期生存戦略の例

Table 6　Examples surviving strategies

戦略	生存期間
0101	89364
1001000100010001	88546
00010001	88504
1101000100010001	88096
1001000100010001	88055
10010001000100011001000100010001	88016
1001000100000001	87539

強い戦略が次々と移り変わる様子が確認できる．01 戦略は第 2 章で述べたように，3 つの優秀な戦略である条件を兼ね備えた戦略である．1001 戦略は 01 戦略よりもノイズに強い性質を持ち，ノイズが存在する環境での 01 戦略同士の対戦の場合，01 戦略よりも高い利得を得ることが出来る．ただし，1001 戦略は裏切り行動主体の戦略には極めて弱い特徴も持っているため，0001 戦略のような，裏切りやすい戦略が相手には勝つことが出来ない．しかし，0001 戦略もまた，01 戦略に弱い．このことから，01 戦略 >1001 戦略 >0001 戦略 >01 戦略 >…という 3 竦みの関係性が構築されていることが分かる．図 3 の結果は，この 3 竦みの関係がもたらした結果と言える．また，図 4 のように，長期間生存している戦略の遺伝子は，01 戦略・1001 戦略・0001 戦略・0000 戦略の 4 つで構成されていることが確認できた．このことから，01 戦略・1001 戦略・0001 戦略の 3 つの戦略が動的環境における有用な戦略なのではないかと考えられる．

5.　おわりに

戦略及び人口分布が刻々と変化する動的な環境において，戦略がどのように変化していくのかを，IPD を用いたシミュレーションを用いて分析を行った．その結果，最も人口の数を大きくした戦略は，アクセルロッドの実験にお

いて優秀な成績を残した 01 戦略（しっぺ返し戦略），ノイズに強い 1001 戦略（pavlov），報復傾向の強い TFT 戦略である 0001 戦略 (R-TFT) であった．また，より戦略を複雑に進化させた長期生存戦略においても，これらの戦略は遺伝子情報として残っていることが確認できた．これらの結果から，動的な環境において有用な戦略は，TFT 戦略，Pavlov 戦略，R-TFT 戦略ではないかと考えられる．今後の課題として，10010001 と 00011001 のような遺伝情報の組み合わせによる戦略の強さの変化や，初期パラメータによる変化などの実験を行い，動的な環境による精密な最適戦略の分析を行っていく．

参考文献

[1] Novak，M．A． and Sigmund，K．: Evolution of indirect reciprocity by image scoring， Nature， Vol．393，pp．573-576 (1998)．
[2] Roberts，G． and Sherratt，T．N．: Development of cooperative relationship through increasing investment，Nature， Vol．394， pp．175-178 (1998)．
[3] Yao，X． and Darwen，P．: How important is your requtation in a multi-agent environment，IEEE-SMC1999，pp．575-580 (1999)．
[4] Robert Axelrod(松田耕治　訳): つきあい方の科学，ミネルヴァ書房 (1998)，原著:The Evolution of Cooperation (1984)．
[5] Kristian Lindgren: Evolutionary Phenomena in Simple Dynamics，Artificial Life II，Addison-Wesley，PP．295-312(1990)．
[6] M．Tanaka-Yamawaki and T．Murakami: Effect of reputation on theformation of cooperative network of prisoners． New Advances in Intelligent Decision Technologies，SCI199(Springer)，pp．615-623 (2009)．
[7] 糸井　良太, 田中　美栄子: 進化型 IPD における共存社会の形成情報処理学会研究報告，Vol.2011-MPS-83 No.2(2011 年 5 月 17 日)．
[8] 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，一般社団法人日本ソフトウェア科学会，コンピュータソフトウェア 14(4)， pp．365-378 (1997)．
[9] 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)．
[10] 田中 美栄子，中武　耕治: 囚人のジレンマ型問題における戦略の進化，宮崎大学工学部紀要，vol30，pp．319-326，(2001)．
[11] 根路銘 もえ子，遠藤 聡志，山田 孝治，宮城 隼夫：繰り返し囚人のジレンマゲームにおける競合共進化戦略の解析に関する考察，電子情報通信学会技術研究報告． CST，コンカレント工学 99(418)， pp．65-70， (1999)．