

## 進化的 IPD における共存社会の形成

糸井良太<sup>†1</sup> 田中美栄子<sup>†1</sup>

進化的 IPD に基づくマルチエージェント人工社会において、戦略を 2 値文字列として遺伝子のように扱い、複写変異と点変異により戦略長をより長く進化させていくと、各エージェントが強調し合う共存社会の出現するケースが見つかった。このような共存社会は長期に渡って続くものではないが、シミュレーションの途中でかなりの期間安定に存在できていた。この期間における戦略を調べると、そのエンタロピーは安定に増大しており、多様性の高い戦略分布を示していた。また、この時の主流となっている戦略の性質を調べると、友好性・寛容性・柔軟性の高い、「心の優しい」戦略が多いことがわかった。

### Formation of cooperative society in the iterated prisoner's dilemma game with evolving genes

RYOTA ITO<sup>†1</sup> and MIEKO TANAKA-YAMAWAKI<sup>†1</sup>

In a multi-agent model of the artificial society based on evolutionary IPD, namely, the iterated prisoner's dilemma game with evolving genes, we construct a model by dealing with prisoner's strategy as binary strings, regarding the strategies as chromosomes, and let their length double and change it by a site mutation. In the study of this model by simulation, we have found a state of cooperative society. Although such a state does not prolong forever, it remains stable for a long time in the course of simulation. The population of strategies in this cooperative state has shown a high level of diversity, which is quantified by means of entropy steadily increasing during the period of cooperation. In this period, the dominant strategies have shown to be "warm-hearted" strategies, having the high level of friendship, tolerance, and flexibility.

## 1. はじめに

囚人のジレンマ (Prisoner's Dilemma: 以下 PD) において、対戦相手との協力関係の構築は非常に難しい問題である。なぜなら PD 自体が敵対する人間や国をモデル化したものであり、個人にとっての最も合理的な選択が裏切り行動となる構造を内包しているためである。しかし、現実社会にはこうした敵対構造の問題が数多く存在する。それに伴い、どのように協力関係を構築するかといった研究が数多くなされてきた<sup>1)2)</sup>。これらの研究の多くは PD 問題に追加要素を加えることによって協力関係の構築を目指しているが、本研究では、複数のエージェントが長期間の協力関係を維持している『共存社会』が、協力に導くための人為的な要素を導入することなく自然に形成されている事例を発見したことに意義がある。実際、PD において『共存社会』の形成は非常に稀な事例である。本研究は、『共存社会』が形成された事例を分析し、形成されなかった事例と比較することで、形成のために必要な条件は何か。また、それによってもたらされる効果は何かを考察をした。

## 2. 囚人のジレンマ

### 2.1 基本概念

PD には一般的に表 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は得点  $R$  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は得点  $S$  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としての最も合理的な選択は裏切り行動になる。自分がそう考える時、相手も同じように考えて裏切りを選択するわけで、この場合の両者の利得はいずれも  $P$  となってしまう、両方で協力し合った時の両者の利得  $R$  より小さくなってしまふ。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。両者で相談して協力

<sup>†1</sup> 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

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

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

$$S+T < 2R, S < P < T < R$$

し合えば良い、と考えるのは相手を無制限に信用できる場合のみであり、むしろ相手がこちらをそこまで信じているのであれば裏切った方が得だ、という人間関係の一面を冷酷にえぐり出しているのが囚人のジレンマだと言える。

## 2.2 繰返し囚人のジレンマ

裏切りを防ぐ手段として、終わりを告げずに PD を繰返し行わせる繰返し囚人のジレンマ (Iterated Prisoners Dilemma: 以下 IPD) がある。繰返し試行を行わせると、過去の行動が次の試行に影響を及ぼす。つまり、早い段階で裏切るとそれ以降相手が協力しなくなり、高い利得が得られない。そのため安易に裏切りが選択できず、協力し合う可能性が生まれる。

## 2.3 進化型 IPD モデル

Lindgren<sup>3)</sup> 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。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<sup>4)</sup>。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、このモデルを使用することで戦略の探索領域を絞ってシミュレーションを行うことができる。

## 3. 進化型 IPD モデルにおける戦略の自動進化

### 3.1 モデル

我々は Lindgren モデルを参考に、繰返し囚人のジレンマをマルチエージェントモデルで扱い、遺伝的アルゴリズムにならって環境により良く適応した戦略を元にしてそこから派生する長い戦略を次々と発生させてみるにより、全戦略空間にわたる探索を避けつつ、

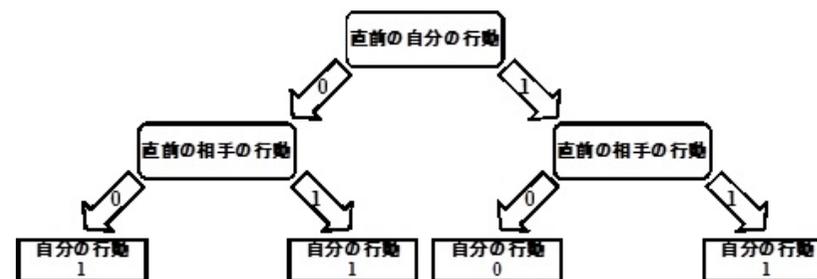


図 1 過去の手と戦略の関係 (m=2 の例)  
Fig. 1 Strategy and history of actions (m=2)

囚人のジレンマで勝てる戦略を自動生成させる研究を行った<sup>4)</sup>。

### 3.2 戦略とノイズ

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 は参照する歴史の長さであり、m=0 が最小で、これは歴史を全く考慮しない場合に相当する。m=1 は直前の相手の手のみを考慮する、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、しつぱ返し戦略 (TFT, TIT for TAT) である。10 はその反対の手を出すので逆しつぱ返し (ATFT, Anti Tit for Tat) 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。1001 はパプロフ戦略と呼ばれる。図 1 に 1101 戦略を例として示した。参照情報数が m となるとき戦略長 L は、

$$L = 2^m \tag{1}$$

となる。ただし、今回の実験では最大戦略は 32 となっている。さらに、エージェントは必ず戦略通りに行動するわけではなく、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 対戦

対戦は戦略単位で行う。また、同じ相手と規定回数続けて PD ゲームを行う。各戦略は自分の戦略に基づいて行動するが、初手は必ず協力行動を行う。また、参照情報は対戦相手が変わるたびに初期化を行っている。以上の試行を総当たりで行う。

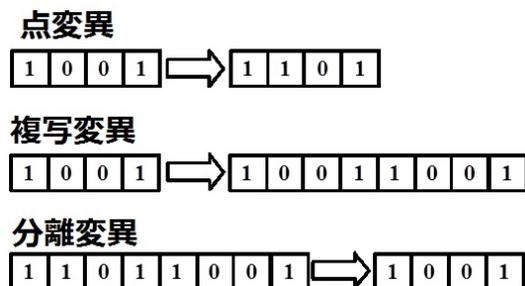


図 2 本稿で考察する 3 種の突然変異 (点変異, 複写変異, 分離変異)

Fig. 2 Three types of mutations used in our model (point mutation, doubling, separation)

### 3.4 人口の変化

遺伝的アルゴリズムに準じた方法で, より良い戦略は模倣を生むと考える. しかし戦略の数に対してエージェント数が多く, 同じ戦略を持つエージェントの行動は同じである. この重複を避けるため, 戦略をエージェントと扱い, 同じ戦略を持つエージェントの数をその戦略の属性として扱うことにより計算時間を短縮する. この考えに基づき次のように考える. 各戦略は人口という要素を持っており, この人口は対戦の結果により変動する. 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 1 試あたり獲得した利得 (以下: 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.

$$X_i(t+1) - X_i(t) = \alpha(S_i(t) - \bar{S}(t))X_i \quad (2)$$

$\alpha$  は増加率を表す. 式 2 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.

### 3.5 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

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.

### 3.6 実験条件

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

表 2 実験条件  
Table 2 Experimental condition

対戦方式	総当たり	増減率	1/100
総人口	1000	点変異	1/9000
対戦回数	1000	複写変異	1/12000
ノイズ発生率	1/100	分離変異	1/12000

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

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. そのため, 裏切りと協力を交互に行った時の平均利得である 2.5 以上の値となることを, 協力関係が構築されているとみなす基準とした.

### 3.7 共存社会の定義

複数の戦略により長期的に協力関係が維持されている状態を「共存社会」と定義する. そのため, 共存社会が形成されているかの判断は

- 全体での平均利得が 2.5 以上の値である
- 複数の戦略が同時に存在している

という 2 条件が長期的に持続していることとする.

### 3.8 分析項目

本研究では次の 3 項目について分析を行っている.

#### (1) 友好性

直前の相手の協力行動に対して協力行動を返す性質と定義している. これは戦略の偶数番目の情報が 1 である数を  $X_e$  としてこれが大きいほど友好的な戦略であるとする. 戦略長を  $L$  としたときの友好度  $F$  は

$$F = \frac{X_e}{L/2} \quad (3)$$

とすると,  $0 \leq F \leq 1$  を満たす指標となる.

(2) 寛容性

直前の相手の裏切り行為を許し協力行動を返す性質と定義している。これは戦略の奇数番目の情報が1である数を  $X_o$  としてこれが大きいほど寛容的な戦略であるとす。戦略長を  $L$  としたときの寛容度  $T$  は

$$T = \frac{X_o}{L/2} \quad (4)$$

とすると、 $0 \leq F \leq 1$  を満たす指標となる。

(3) 柔軟性

行動に偏りが少ない性質と定義している。戦略は  $[00] \cdot [01] \cdot [10] \cdot [11]$  の4つの戦略パターンの組み合わせで構成されている。その戦略パターンの発生確率から平均情報量を求める。戦略パターンは4通りであるので最大値は約1.38となり、値が最大値に近いほど柔軟性にとんだ戦略とする。各戦略パターンの発生確率をそれぞれ  $P_{00}$ ,  $P_{01}$ ,  $P_{10}$ ,  $P_{11}$  としたときの柔軟度  $E$  は

$$E = -(P_{00} \ln P_{00} + P_{01} \ln P_{01} + P_{10} \ln P_{10} + P_{11} \ln P_{11}) \quad (5)$$

となる。柔軟性の検証は協力行動後と裏切り行動後の2通りで検証を行う。

4. 実験結果

4.1 共存社会

4.1.1 主要戦略の複雑化と共存社会の形成

戦略別の人口推移 (Population) と1試合あたりの全体での平均利得 (Average Score) を図3に示す。但し、図を煩雑にしないため戦略数150未満の戦略を非表示に設定している。0~20000世代間は戦略長2~4の単純な戦略が主流だが、25000~30000世代間は戦略長16の様な、より長く、より複雑な戦略が台頭しており、戦略長2~4のような単純な戦略は淘汰されている。また、35000~60000世代間では戦略長32の非常に複雑な戦略が主要な戦略になっており、世代が進むにつれて戦略が複雑化している。また、25000~60000世代の長期間にわたって平均利得が基準値を超え、かつ、複数の戦略が同時に存在している。つまり、この期間において『共存社会』が形成されている事が分かる。

4.1.2 多様性の増加

平均利得と戦略別人口の平均情報量 (Entropy) の比較を図4に示す。共存社会が形成されてからのエントロピーは、大きなブレも無く安定して増加している。このことから、人口の大きな推移が無くなり、今まで少数であった戦略の人口も増加することで、各戦略間で人

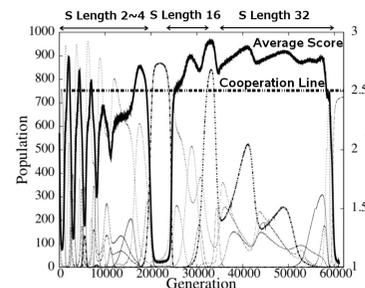


図3 共存社会の形成に成功した例  
 Fig.3 Successful example in making a cooperative society

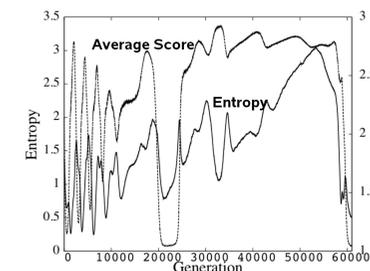


図4 戦略の安定した多様化  
 Fig.4 Stable diversification of strategy

口差が少なくなっていることが分かる。つまり、戦略の多様化が起きている。

4.1.3 共存社会を構成する戦略の特徴

(1) 友好性

共存社会を構成する主要な戦略の友好性を図5に示す。60~70%を中心に多くの戦略が高い値で分布している。つまり、友好性の高い戦略により共存社会は形成されている。

(2) 寛容性

共存社会を形成する主要な戦略の寛容性を図5に示す。40~50%に殆どの戦略が集中している。つまり、半分の確率で相手の裏切りを許容する戦略で占められている。

(3) 柔軟性

図6は戦略別の戦略パターン発生確率の柔軟性をヒストグラムで示している。図6より、ほぼ全ての戦略が偏りの少ない戦略であることが確認できる。つまり、各戦略とも協力、裏切り関係なく戦略パターンの発生確率が等分布している。

4.2 非共存社会

4.2.1 共存社会の非形成

共存社会が形成されなかった時の戦略別人口の推移を図7に示す。但し、最大人口数が100未満の戦略は非表示にしている。戦略長4の戦略が台頭し始める10000世代から様々な戦略長の戦略が出現し台頭するが、その戦略長は図3とは異なり安定せず、単純な戦略と複雑な戦略が入り乱れて存在している。また、平均利得が2.5以上で安定している期間が存在するが、その期間中は1つの戦略に人口が集中し他の戦略が淘汰されているため、共存社

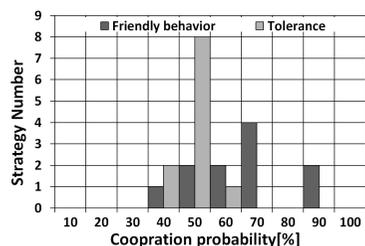


図 5 共存社会の友好性と寛容性

Fig. 5 Friendly behavior and tolerance of a cooperative society

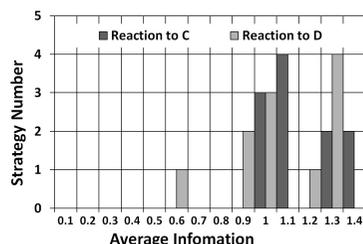


図 6 共存社会の柔軟性

Fig. 6 Flexibility of a cooperative society

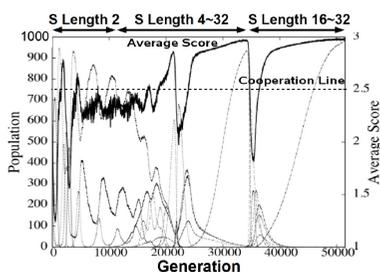


図 7 共存社会の形成に失敗した例

Fig. 7 Unsuccessful example in making a cooperative society

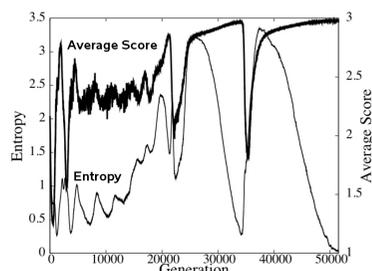


図 8 多様性の減少

Fig. 8 Decrease of diversity

会が形成されていると言い難い。

#### 4.2.2 多様性の減少

平均利得と戦略別人口の平均情報量の比較を図 8 に示す。平均情報量は一貫して安定せず増加と減少を繰り返している。これは、共存社会が形成されないため戦略人口の移り変わりが激しく、戦略の絶滅と誕生が頻繁に起こっているためだと考えられる。

#### 4.2.3 非共存社会を構成する戦略の特徴

##### (1) 友好性

分析結果を図 9 に示す。主要戦略の半数以上が 50% 以下の友好度であり、協力をし返す戦略が少ない。友好度が低いと協力を続けて行うことができないため、共存社会の形成の障害になる。

##### (2) 寛容性

分析結果を図 9 に示す。多くの戦略が低い値に偏っており、相手の裏切りには瞬時に裏切りで報復し返す戦略が多数存在している。つまり、ノイズによる失敗も許さない戦略が多く、それが共存社会の形成の妨害になったと考えられる。

##### (3) 柔軟性

戦略別の柔軟性のヒストグラムを図 10 に示す。平均情報量が 1 以下の値の戦略が多数存在している。このことから、行動に偏りのある戦略が多いことが確認できる。特に各戦略の裏切り行動後の [00] 戦略を行う確率が、全体平均で 50% 以上の確率で存在していた。つまり、裏切り行動を続けて行う戦略が多く存在していたことが確認できた。

#### 4.3 共存社会と非共存社会の比較

##### 4.3.1 友好性の比較

図 11 に共存社会と非共存社会との友好性の比較を示す。共存社会が 60% を中心に分布しているのに対し、非共存社会は疎らに分布している。つまり、共存社会が友好的な戦略で構成されているのに対し、非共存社会は友好的な戦略とそうでない戦略が同時に存在している。

##### 4.3.2 寛容性の比較

図 12 に共存社会と非共存社会との寛容性の比較を示す。共存社会が 40~50% に集中しているのに対し、非共存社会は値の低い方に偏って分布している。非共存社会では、ノイズによる行動の失敗も許さない戦略が多数存在していることが確認できる。

##### 4.3.3 柔軟性の比較

図 13 と図 14 に、共存社会と非共存社会の協力的行動と裏切り行動後の行動パターンの平

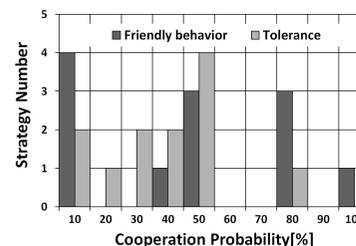


図 9 非共存社会の友好性と寛容性

Fig. 9 Friendly behavior and tolerance of non cooperative society

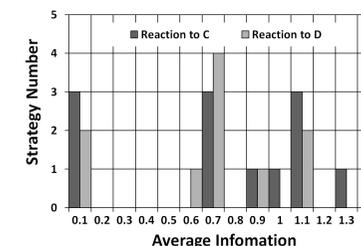


図 10 非共存社会の柔軟性

Fig. 10 Flexibility of non cooperative society

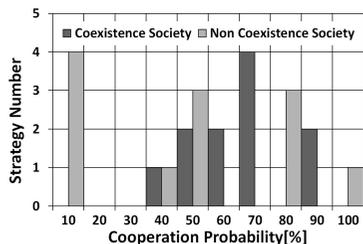


図 11 友好性の比較

Fig. 11 Comparison of friendly behavior

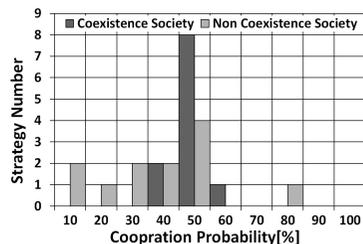


図 12 寛容性の比較

Fig. 12 Comparison of tolerance

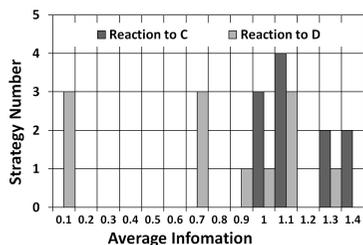


図 13 協力行動後の行動の偏り

Fig. 13 Comparison of actions after cooperation

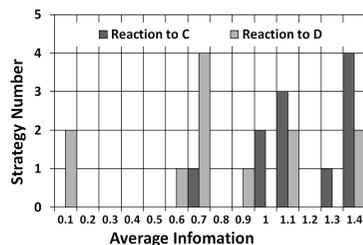


図 14 裏切り行動後の行動の偏り

Fig. 14 Comparison of actions after defection

均情報量の比較を示す．共存社会を構成する多くの戦略が、行動にかかわらず 1 以上の値に集中しており、偏りの少ない複雑な戦略で占められていることが分かる．それに対して非共存社会を構成する多くの戦略が、協力行動、裏切り行動ともに値が低く、行動の偏りが大きい単純な戦略で占められていることが分かる．また、各行動パターンの発生確率を比較しても、協力社会の戦略に比べ非共存社会の戦略は必ず裏切る戦略パターンである [00] の確率が高い値であった．

## 5. 結 果

共存社会を構成する戦略とそうでない戦略を比較した結果、非共存社会の主要な戦略が裏切り行動に偏った戦略であるのに対し、共存社会を構成する戦略は友好性・寛容性・柔軟性の高い「心の優しい」戦略であることが分かった．共存社会と非共存社会での多様性の変化を

比較すると、非共存社会の多様性が安定せず常に増加・減少を繰り返していたことに対して、共存社会の多様性は安定して増加していた．また、囚人のジレンマにおいて相手の直前の行動を返すしっぺ返し戦略が最も効率が良いとされている<sup>5)</sup>が、今回の実験では必ずしもしっぺ返し戦略が勝ち残っていたわけではなく、しっぺ返し戦略が絶滅する場合も存在した．2004 年に開かれた大会においても、しっぺ返し戦略が敗れた戦略は他の戦略と競合する戦略<sup>6)</sup>であった．つまり共存社会においていしっぺ返し戦略は必ずしも最良の戦略ではないと考えられる．

## 6. おわりに

囚人のジレンマゲームにおいて「心の優しい」戦略が勝ち残ることは難しい．しかし、本研究の事例では「心の優しい」戦略が勝ち残り、共存社会を形成している．このような事例が発生した要因として考えられるのが戦略の大量淘汰である．裏切り主体戦略が淘汰されなければ「心の優しい」戦略が勝ち残ることは難しい．また「心の優しい」戦略にはある程度の戦略長が必要であるため戦略の複雑化は必須であるため、単純な戦略よりも優位にならなくてはならない．事実、図 3 において、主要戦略が入れ替わる節目には、大きく Entropy が下がっている時期が存在する．この時期に裏切り主体戦略や単純な戦略が淘汰されるかどうか共存社会が形成されるかの重要な節目でないかと考えられる．今後の課題としては、本研究の結果をもとに共存社会形成のためのより詳しい分析を行っていく．

## 参 考 文 献

- 1) 小野 歩, et.al.: 友人ネットワークの発展により IPD に出現するスケールフリー構造, 第 21 回自律分散システム・シンポジウム資料 (SICE2009), pp.187-190 (2009).
- 2) 村上 拓, et.al.: 合意形成による市場の安定化を目指した評判付き囚人のジレンマモデルの検討, 第 21 回自律分散システム・シンポジウム資料 (SICE2009), pp.187-190 (2009).
- 3) Lindgre, K.: Evolutionary Phenomena in Simple Dynamics, *Artificial Life II*, pp. 295-312 (1990).
- 4) 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, Vol.30, pp.319-326 (2001).
- 5) 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).
- 6) 湯田賢司, 多々良和臣: 戦略ゲーム『繰り返し囚人のジレンマ』記念大会開催 (2004). <http://wiredvision.jp/archives/200410/2004101503.html>.