協調の進化
繰り返し囚人のディレンマ・ゲームの
コンピューター・シミュレーション
Evolution of Cooperation:
A Computer Simulation of Iterated Prisoner's Dilemma

高増　明・服部純典
Akira Takamasu and Suminori Hattori
大阪産業大学経済学部

1. はじめに

囚人のディレンマ・ゲーム（Prisoner's Dilemma Game）は、現在、自然科学、社会科学の両分野において注目されている。このゲームは、協調（Cooperate）と裏切り（Defect）という二つの戦略をもつ二人のプレイヤーが対戦するとき、それぞれが自己の利益のために相手を裏切る戦略を選択し、結果として、両者が協調戦略をとったときと比較して、どちらにとっても不利な結果が生じるというものである。このような状況は、自然界や社会に幅広く存在するが、現実には多くの場合、協調的な戦略が選択されていると考えられる。したがって、どうしてそうなるのかが様々なかたちで議論されてきた。

そうしたなかで、ゲームが 1 回限りなら、たしかに裏切りが合理的な選択となるが、ゲームが無限に繰り返されるときには、協調戦略が採用される可能性が存在することが明らかになってきた。これは、直観的な表現を使えば、相手を裏切ることによって短期的には大きな利益を得るが、そのことによって他のプレイヤーの信用を失い、将来的には小さな利益しか得られなくなるからである。

このような状況、すなわち囚人のディレンマ・ゲームが繰り返されるゲームは、繰り返し囚人のディレンマ・ゲーム（Iterated Prisoner's Dilemma: IPD）と呼ばれる。IPD は、自然科学や社会科学における様々な現象と関連づけて研究が行われている。自然科学、たとえば数理生態学では、どのようにして生態学的なコミュニティーが形成されていくのかを人工生命（Artificial Life）[1]や遺伝的アルゴリズム（Genetic Algorithms）[2]の分析方法を使って研究している。その基本的な考え方は、様々な戦略が対戦するなかで、より環境に適応した戦略が生き残り、さらに環境に適応するように戦略が進化していくというものである。そうした進化と適者生存のプロセスを通して、協調あるいはコミュニティーが出現してくると考える。

一方、社会科学では、ゲーム理論の研究者によって、無限繰り返し囚人のディレンマ・ゲームでは、協調を選択することが合理的な選択になることが証明されている（「フォーク定理」）。しかし、最近では、生物学者の Maynard Smith[3]らによってはじめられた進化ゲーム（Evolutionary Games）のアプ

---

[1] 人工生命については、Langton(1989), (1996)所収の論文などを参照してもらいたい。
[2] 遺伝的アルゴリズムの概要については Goldberg(1989)が明解である。邦文文献としては、伊庭(1994)がある。
[3] Maynard Smith, J. and G. A. Price (1973)が進化ゲームについての最初の論文である。また Maynard Smith(1982)も参照してもらいたい。

ローチを使い、プレイヤーの合理性が限定されたものである（限定合理性：bounded ratinality）と想定して分析が行われている。そして、社会秩序、社会ルール、社会制度や慣習が、進化と適応のプロセスのなかで次第に形成されていくことを示そうとしている。

繰り返しゲームの分析では、最適な戦略を数学的に導き出すことがむずかしいため、しばしばコンピューター・シミュレーションが用いられる。このようなコンピューター・シミュレーションのアプローチは、Axelrod(1981)によって始められた。Axelrod は、数学者、政治学者、社会学者、経済学者などに、繰り返し囚人のディレンマ・ゲームを行うコンピューター・プログラムを作ってもらい、そのプログラムを対戦させるトーナメントを行った。トーナメントでは、ひとつのプログラムが、別のプログラムと 200 回ずつ対戦し、様々なプログラムと対戦した総得点によって順位をつける。その結果は、「第 1 手は協調を選択するが、それ以降は前回に相手が選んだのと同じ手を選択する」という単純な戦略が優勝した。この戦略は、「しっぺ返し戦略」（Tit For Tat: TFT）戦略とよばれる。TFT は、さらに、トーナメントの結果を考慮して参加者がプログラムを改良し 62 の戦略が集まった第 2 回のトーナメントにも優勝している。

コンピューター・シミュレーションによる IPD の分析は、その後多くの研究者によって発展させられていったが、最近では、Lindgren-Nordahl(1991)、(1993)、(1996)が、人工生命という分野において、戦略の進化、ノイズの存在などを考慮にいれて詳細に検討している。ノイズ、すなわちプレイヤーが本来選択するはずではない戦略を誤って選択してしまう可能性を考慮にいれるときには、必ずしも TFT が最強の戦略にはならないからである。

この論文でも、これらの先行する業績、ゲーム理論、進化ゲーム、遺伝的アルゴリズムによる IPD の分析を紹介するとともに、Lindgren-Nordahl を基礎として、コンピューター・シミュレーションを行いたい。その際、Lindgren-Nordahl では手続きが若干不明確な点を厳密に議論し、また Lindgren-Nordah では議論されていない利得行列の数値の違いを検討する。

論文の内容を簡単に紹介しておこう。第 2 節では囚人のディレンマ・ゲームについて簡単に説明する。第 3 節では、繰り返し囚人のディレンマ・ゲームについて、ゲーム理論がそのナッシュ均衡をどのように分析しているのかを紹介する。第 4 節では進化ゲーム、第 5 節では遺伝的アルゴリズムの方法論とそれが囚人のディレンマ・ゲームとどのように関係しているのかを説明する。第 6 節では、コンピューター・シミュレーションを行う。メモリーが 1 でノイズが存在しないケースについて、戦略の初期の個体数がどのようにそれぞれの戦略の生き残りに影響するのかを考察する。続いて、利得行列の数値を変更し、戦略の生き残りがどのように変化するのかを調べる。最後に第 7 節で、このような研究の今後の発展の可能性について簡単に展望する。

2. 囚人のディレンマ

囚人のディレンマとは、つぎのような状況のことを指す。

囚人のディレンマ

裁判をひかえた二人の囚人がいる。二人は共同で重大な犯罪を犯したことがわかっているが、検察

側はそれを立証するのに十分な証拠をもっていない。一方、軽い犯罪の証拠は揃っている。そこで、囚人はそれぞれ別々に尋問され自白を促されている。二人がともに自白するならば、かれらは、重大な犯罪を犯した罪で懲役 10 年となる。両者が完全に黙秘すれば、軽い犯罪による 2 年の刑になるだけである。もし一人が自白し、もう一人が黙秘した場合には、自白した者は検察との取引によって釈放され、他方は 20 年の刑になる。このような状況のとき二人の囚人はどのような行動をとるだろうか。このとき二人の行動、すなわち黙秘、自白と二人の刑期は以下の表 1 のようになっている。表 1 でカッコの最初の数字は囚人 1 の刑期を 2 番目の数字は囚人 2 の刑期を表している。

		囚　人2	
		黙　秘	**自　白**
囚　人1	**黙　秘**	(2,　2)	(20,　0)
	自　白	(0,　20)	(10,　10)

表1　囚人のディレンマ

このような状況におかれたときに囚人 1 がどのように行動するのかを考えてみよう。囚人 1 は、もし囚人 2 が黙秘するときには、黙秘すれば 2 年の刑、自白すれば釈放されるから自白するのが有利である。一方、囚人 2 が自白するときには、黙秘すれば 20 年の刑、自白すれば 10 年の刑なので、やはり自白するのが有利である。このように囚人 1 にとっては囚人 2 がどのような行動をとろうと自白するのが有利になる[4]。このことは、囚人 2 にとっても同様であるから、結局どちらの囚人も自白し、結果として、どちらも 10 年の刑に服することになる。しかしながら、二人が単純に相手のことを考えて黙秘という行動を選択すれば、どちらも 2 年の刑ですむわけである。このように二人が合理的に行動した結果、どちらにとっても不利な結果となるケースを囚人のディレンマと呼ぶ。

この囚人のディレンマが注目されるのは、このような状況が、社会や自然界において広範に存在しているからである。たとえば、囚人のかわりに、二つの企業を想定し、黙秘を他企業と協調的な戦略、自白を他企業に対して裏切り的な戦略と置き換えてみよう。一方の企業が協調的な戦略をとっているときに、もう片方の企業が自社の利益だけを考えて、安値で販売したり、過大な広告費を投入するという裏切り的な戦略をとるときには、裏切り的な手をとった企業は大きな利益を得ることができるだろう。しかしながら、他企業も同様に考えて、同じ裏切り的な手をとれば、結局はどちらの企業も両社が協調したときよりも低い利益しか得ることができないことになってしまう。したがって、二つの企業が競争するケースもまた囚人のディレンマと考えることができる。また囚人を二つの政府（たとえば日本とアメリカ）、商店とその顧客、友人同士、恋人同士などに置き換えても、同じような状況が想定できるだろう。

ここで、後の議論に都合のいいように、囚人のディレンマの状況を刑期のかわりに、各プレイヤー

[4] このような戦略を支配戦略（dominant strategy）という。

の利得を示すかたちで、つぎのように再定義しておこう。

プレイヤー 2

		協　調（C）	裏切り（D）
プレイヤー 1	協　調（C）	(R, R)	(S, T)
	裏切り（D）	(T, S)	(P, P)

表 2　囚人のディレンマの利得行列による定義

R はお互いに協調したときの報酬である。T は協調しようとする相手に対して裏切ろうとする誘惑である。このとき裏切られたほうは S だけしか報酬を手に入れることができない。また P は両者が裏切ったときの利得である。囚人のディレンマは、$T > R > P > S$、$2R > T + S$[5]という不等式によって特徴づけられる。この不等式を満たす T、R、P、S のうち、一般によく使われる R=3、T=5、S=0、P=1 というものを使おう。

プレイヤー 2

		協　調（C）	裏切り（D）
プレイヤー 1	協　調（C）	(3, 3)	(0, 5)
	裏切り（D）	(5, 0)	(1, 1)

表 3　囚人のディレンマの利得行列による定義（数値を与えたもの）

ただし、利得の大きさによって結果が変化すると考えられるため、コンピューター・シミュレーションでは、他の数値も使うことにする。

3. 繰り返し囚人のディレンマゲームとナッシュ均衡

囚人のディレンマのケースにおいて、どちらにとっても不利になる結果を避けるには、どうしたらいいのだろうか。このような状況が 1 回かぎりであれば、確かに裏切り的な手を採用することが合理的な行動である。しかしながら、このような状況が繰り返し起こるとすれば、一度裏切り戦略をとると、それ以後は他人に警戒され、相手も裏切り的な手しかとらなくなるだろう。したがって、囚人のディレンマが繰り返されるケースでは、協調的な手をとることが合理的になる可能性が出てくる。ゲーム理論では、このようなケースについて、「フォーク定理（folk 定理）」と呼ばれる定理が証明されていて、協調的な戦略を両者がとることが、ナッシュ均衡になる可能性が明らかにされている。フ

[5] この条件は、二人のプレイヤーが裏切りと協調を交互に繰り返したときに得られる利得が、ずっと協調した場合を下回ることを意味している。

ォーク定理とは、つぎのような定理である。

フォーク定理

　個々のプレイヤーが将来のことを重要視するならば、すなわち将来の利得の割引率が十分に 1 に近いときには、個々のプレイヤーが得られる最小の利得（ミニマックス利得）を超えるどんな利得もナッシュ均衡となるような割引率が存在する。

定理の証明についてはゲーム理論のテキストを参照してもらいたい[6]が、この定理をもうすこし、わかりやすく言うと、無限回繰り返しゲームでは、どちらも「協調」を選択するという「協調解」を含む様々な解がナッシュ均衡となるような割引率が存在するということである。たとえば、以下のような戦略もナッシュ均衡となることがわかる。

> 相手が「協調」をとり続けるかぎり「協調」を選択するが、ひとたび相手が裏切ったら、それ以後は裏切り続ける。

この戦略はトリガー戦略と呼ばれるが、このような戦略を二人のプレイヤーがとることは、ナッシュ均衡である。なぜなら、たとえば割引率を 0.9 とするとき、このような戦略をとったときの将来の利得の割引現在価値は 30 である。しかし、そうでなく「エゴイスト戦略」を一度でも選択すると現在は 5 だけの利益を得られる（短期的利益）が、将来の利得は 1 になるからである。利得の割引現在価値は 14 になる。こうしてみると確かに、囚人のディレンマが無限解繰り返されるとしたら、「協調」が均衡となりえることがわかるだろう。しかしながら、問題は、ナッシュ均衡がひとつではないということである[7]。たとえば以下のような戦略もナッシュ均衡となる。

> プレイヤー1 は、プレイヤー2 が協調をとり続けているかぎり、協調と裏切りを交互に実行するが、プレイヤー2 がひとたび裏切りを実行したら、それ以後は裏切りをとり続ける。
>
> プレイヤー2 はプレイヤー1 が協調と裏切りを交互にとり続けているかぎり協調を選択するが、プレイヤー1 がその戦略からはずれたときには、それ以後は裏切りをとり続ける。

このような戦略をとるとき、プレイヤー1 が協調から出発するときには、プレイヤー1 は、3 ポイントと 5 ポイントを交互に獲得するから、その将来利得の割引現在価値は 36.84、プレイヤー2 は 3 ポイントと 1 ポイントを交互に獲得するから 20.53 の将来利得の割引現在価値を得ることになる。プレイ

---

[6] folk 定理については、たとえば Fudenberg and Tirole(1992, pp.150-155)、Binmore(1992, pp.369-377)を参照せよ。この定理が folk 定理と呼ばれているのは、この定理が古くから知られていて、しかも誰の貢献によるものか確かではない、すなわち folklore だからである。

[7] ナッシュ均衡の問題点については、Kreps(1990, chap.5)などを参照してほしい。

ヤー1 がプレイヤー2 より将来利得が大きくなっているにもかかわらず、これがナッシュ均衡になっているのは、この均衡から乖離するときには、プレイヤー1 もプレイヤー2 も将来の利得の割引現在価値が小さくなるからである。

このように、囚人のディレンマゲームを無限回繰り返すことによって、確かに協調がナッシュ均衡解として得られることになるが、均衡はそのほかにも多数存在し、どの均衡が選ばれるのかは明らかではない。

しかしながら、ゲームのプレイヤーに「神のような」合理性を仮定し、ナッシュ均衡によって分析を行っていくゲーム理論のアプローチはかなり極端なものであることはゲーム理論に批判的な多くの論者が指摘していることである。したがって、つぎに、プレイヤーのとる戦略は変化しないと仮定し、いろいろな戦略のなかでより大きな利得をあげた戦略が生き残っていくという進化ゲームのアプローチ、戦略が進化していく遺伝的アルゴリズムを紹介していきたい。

4. 進化ゲーム

進化ゲームでは、個々のプレイヤー（個体）は複数の選択肢のなかから行動を選択し、その行動の結果として各プレイヤーの利得あるいは適応度（fitness）が決定される。時間が経過するにしたがって、適応度（利得）のより高い戦略を取るプレイヤーの数は増加し、逆に適応度の低い戦略を取るプレイヤーの数は減少していく。そのような様々な戦略を取るプレイヤーの数が増加するのか減少するのか、あるいは定常状態に収束していくのかといったダイナミクスを考察するのが進化ゲームのアプローチである。

進化ゲームは、生物学において John Maynard Smith らによって始められ、数学者によってより厳密なかたちに発展させられていったが、そこでの均衡概念は、「進化的に安定な戦略（ESS: Evolutionary Stable Strategy）」である[8]。ESS はつぎのように定義される。

進化的に安定な戦略（ESS: Evolutionary Stable Strategy**）**
もし集団の全員がその戦略を採用していれば突然変異によって侵入してくる他のどんな戦略をとる個体でも、自然選択の力によって排除できるものをいう。

Axelrod は、ゲームがつぎの回で終了する可能性が十分低いとき（ゲームが十分長く続くとき）には、TFT（しっぺ返し戦略）が ESS になることを証明した。証明の概要はつぎのようなものである。TFT に勝つ可能性をもつ戦略としては、協調を繰り返す戦略（CCCC・・・・・・）、裏切りを繰り返す戦略（DDDD・・・・・・）、裏切りと協調を繰り返す戦略（DCDC・・・・・・）だけを考えれば十分である。なぜなら、TFT は 1 回分の記憶（相手が前回どの手をとったか）しかもたないし、ゲームは将来のどの時点でも、ゲームがそれから先続く回数の期待値は同じだからである。

このとき、ある戦略が TFT に対する最適な戦略だとし、もし、それが最初に C をとるとしたら、つ

[8] この節の議論については、Maynard Smith (1982)を基礎にしている。

ぎも C をとらなければならない。なぜなら TFT は、はじめに C をとることがわかっているわけだから、それに対して C をとることが最適なら、TFT は、つぎも C をとるから、最適な戦略は再び C をとらなければならないからである。したがって最適な戦略は C をとり続けるだろう。

逆に、最適な戦略が、もし最初に D をとったとしよう。このとき TFT はつぎに D をとるから、それに対して D をとるとしたら、その戦略は D をずっととり続けなければならない。また最初に D をとって、つぎに C をとるとしたら、つぎの回に TFT は C をとるから、最適な戦略は D をとり、DCDC‥‥……となるからである。したがって、この三つが TFT に勝つ戦略の候補となりえるだろう。

TFT が TFT 自身と戦うときには、その将来利得は、ゲームがつぎの回も継続する確率を w とすると、

$$R + wR + w^2 R + \cdots = \frac{R}{1-w}$$

となる。これは、R=3、w=0.9 とすると、30 になる。また CCCC‥‥…が TFT と対戦するときも利得の期待値は同じだから、CCCC‥‥…は侵入できない。つぎに DCDC‥‥…が、TFT と対戦するときには、その利得の期待値は、

$$T + wS + w^2 T + w^3 S + \cdots = \frac{T + wS}{1-w^2}$$

となり、これは、T=5、S=0、w=0.9 とすると 26.3 になる。したがって、DCDC も侵入することはできない。最後に DDDD‥‥…が TFT と対戦するときには、

$$T + wP + w^2 P + \cdots = T + \frac{wP}{1-w}$$

となり、これは T=5、P=1、w=0.9 とすると、14 になり、やはり侵入できない。したがって、w が十分 1 に近いときには、TFT は ESS になることがわかる。

5. 遺伝的アルゴリズム

遺伝的アルゴリズム（GA: Genetic Algoruthms）とは、与えられた問題の最適解をみつけるためのひとつのアルゴリズムである。最適解をみつけるアルゴリズムとしては、ニュートン法など、様々な方法が知られているが、これまでの方法は、たとえば目的関数が二つ以上の「山」をもっているときに間違った「山」に登ってしまう可能性が存在する、関数が連続ではないときに解を見つけられない、すべての値についてチェックしていくために時間がかかりすぎる、などの問題点をもつことが知られている。

GA は、これまでの最適化アルゴリズムよりも効率的に最適解を発見することができると言われている。そのアイディアを一言でいえば、生物進化の適者生存の原理（ダーウィニズム）を最適化問題に応用したものということができる。つまり、より環境に適合している個体が生き残り、その個体が、さらに遺伝子の交叉や突然変異によって、より環境に適合した個体を産み出していくという生物進化の原理を一般の最適化問題のアルゴリズムとして利用したのが GA である。したがって、遺伝的アルゴリズム（GA）という名称がつけられているのである。

一般の最適化問題について GA を説明すると、まず変数をある一定の方法でコード化する。これは

生物学でいえば、生物のそれぞれの個体の遺伝子を特定するということである。つぎにある複数の変数の値（生物でいえば個体群）から出発して、その変数（コード）の値についての目的関数の値を調べてみる。これは生物学で言えば個々の生物の環境に対する適合度（fitness）を調べてやるということである。つぎに、その適合度に応じて、つぎの段階の変数（生物でいえば次世代の個体）が確率的に決定（産み出）される。高い適合度をもつ遺伝子型は数多く生き残り、そうでない型はあまり残らないように次世代が産み出されていくということにしてやる。そうして産み出された変数のコードが、遺伝子と同じように、確率的に交叉や突然変異を起こすとしよう。こうして誕生した新たな変数（個体）は、その目的関数の値（適応度）が計算され、以下同じようにして、適合度に応じて次の世代が産み出されていく。このようにして、最適値を発見していこうというのが、遺伝的アルゴリズムである。

この遺伝的アルゴリズムを使って、Axelrod(1985)は、TFT に勝つ戦略を創り出そうとした。Axelrod の方法は以下のようなものである[9]。

①各戦略は過去 3 回の対戦を記憶している、すなわちメモリーが 3 であるとする。それぞれの対戦について、自分と相手がとった手に応じて以下の表の右端の列のようにコードをつけることにする。

プレイヤー1（自分）	プレイヤー2（相手）	記述（自己の得点）	コ　ー　ド
C	C	R	0
D	C	T	1
C	D	S	2
D	D	P	3

表４　過去の結果のコード化

② それぞれの対戦について 4 通りの結果があるわけだから、過去 3 回の対戦では、64 通りの結果が存在する。各戦略（プレイヤー）は、それぞれの結果に対して、協調 1 か裏切り 0 の二つの選択を行うとする。たとえば、過去 3 回の結果が、1 回前が R、2 回前が S、3 回前が T だとする。この結果を R/S/T と記述し、さらにコードで表すと 0/2/1 となる。このコードを 10 進法にすると

$$16\times 0+4\times 2+1=9$$

となる。したがって、この結果に対したプレイヤーが選択する手は 9 ビットのところに 0 あるいは 1 で示すことにする。このようにして、過去 3 回の結果に対する戦略をすべて書いていくと以下のような 64 ビットのコードになる。たとえば

0011101011101101000110110101101101001111000010101110100100000

という戦略は、

1 回前の結果	2 回前の結果	3 回前の結果	選択する戦略

[9] この節の議論については、伊庭(1994, pp41-49)、Pounstone(1992)を参考にしている。

R	R	R	0
R	R	T	0
R	R	S	1
R	R	P	1
R	T	R	1
･･･	･･･	･･･	･･･
P	P	P	0

表 5　戦略コード

というものである。

③　さらに、はじめの 3 回に選択する自分と相手の手（6 ビット）を加えて、総計で 70 ビットのコードをつくる。

Axelrod は、コンピューター・トーナメントから 8 つのプログラムを選んで、20 個のこうしたメモリー3 をもつコードと対戦させた。それぞれは 151 回戦い、その総得点が多いか少ないかに応じて次期の戦略の数が変化し、さらにそれが、遺伝子（コード）の交叉や突然変異によって変化していくと考える。そして、そのようなプロセスのなかから、TFT に勝つ戦略を産み出すことに、Axelrod は成功したわけである。

6. コンピューター・シミュレーション

Axelrod の遺伝的アルゴリズムでは、TFT に勝つ戦略を発見することに力点が置かれている。また「繰り返し囚人のディレンマ」に関するトーナメントでも、どの戦略が強いかということが焦点になっている。しかし、社会科学においては、社会のなかで様々な戦略が出会い対戦するときには、それぞれの戦略をとるプレイヤーの数はどのように変化し、社会にどのような秩序が形成されていくのかという点がより重要である。ここでは、コンピューター・シミュレーションによって、いくつかの戦略が対戦するときに、どの戦略が生き残っていくのかを考察していこう。

6. 1 シミュレーション 1

まず、表 3 のケースについてシミュレーションを行ってみよう。社会には、たくさんのプレイヤーが存在するとし、ランダムに出会って対戦を行うとしよう。プレイヤーがもっているメモリー（記憶）は 1 だとする。このとき、各プレイヤーは前回対戦相手のプレイヤーが選択した手（協調、裏切り）だけを記憶していて、その手に応じて、今回の手を選択するとしよう。このとき対戦相手の前回選択した手は二通り、それに対する手は二通りあるから、4 通りの戦略が存在する。遺伝的アルゴリズムにしたがって各戦略を遺伝子のタイプとして表現することにしよう。1 を協調を、0 を裏切りとすると、記憶が 1 なら、前回の対戦で相手が選択した手に応じて、00、01、10、11 という四つの戦略が存在する。このうち最初の数字は相手が前回 0、すなわち裏切りを選択したときに、その戦略がどの手

を選択するか、2 番目の数字は前回相手が 1、すなわち協調を選択したときに、今回その戦略がどの手を選択するのかを示している。したがって各戦略は、つぎのような戦略と考えることができる。

00：ALLD　相手が前回どのような手をとっても裏切る。
01：TFT　最初は協調を選択する。2 回目からは、相手が前回裏切れば裏切り、相手が前回協調すれば協調する。
10：ATFT　最初は裏切りを選択する。2 回目からは、相手が前回裏切れば協調し、協調すれば裏切る。
11：ALLC　相手が前回どのような手をとっても協調する。

この四つの戦略が、ランダムに出会って対戦するとしよう。そして一定の回数ランダムに対戦が行われた後に、それぞれの個体が獲得した利得に応じて、次の世代の個体数が決まるとしよう。このときには、どの戦略が生き残るのだろうか。

つぎのようなシミュレーションを行った。

シミュレーション

① 00、01、10、11 という四つの遺伝子タイプの初期の個体数を与えてやる。合計は 100 にする。
② 100 個の個体について、第 1 回めの対戦の対戦相手を乱数によって決定する。
③ 第 1 回めに、各遺伝子タイプは、00 は裏切り（0）、01 は協調（1）、10 は裏切り（0）、11 は協調（1）を選択する。
④ 第 1 回めの対戦が行われ、それぞれの利得が決定される。第 1 回めで、各個体が選択した戦略は、その個体の前期のヒストリーになる。
⑤ 第 2 回めの対戦の対戦相手を乱数によって決定する。第 2 回めの対戦において、各遺伝子タイプが選択する戦略は対戦相手のヒストリーによって決定される。たとえば 01 という遺伝子タイプであれば、前期に対戦相手が 0 を選択した場合には 0、1 を選択した場合には 1 を選択する。
⑥ このようにして 30 回の対戦が行われ、各個体についての利得の合計を遺伝子タイプごとに集計し、個体数で割って平均を求める。また総平均も求める。遺伝子タイプの利得の平均が総平均より大きい遺伝子タイプは、その個体数がその差に比例して増加し、逆に小さい遺伝子タイプはその差に比例して減少する。
⑦ ②に戻って対戦を行う。②～⑥を繰り返して、各遺伝子タイプの増減をみる。

シミュレーション 1 の結果

このシミュレーションからつぎのような結果が得られた。

・初期値を 00 が 25、01 が 25、10 が 25、11 が 25 と四つの戦略をとる個体数を全く同一にしてやる。このとき、スタートすると、すぐに裏切り型 00 が増えていく。まず正直な 11 が死滅し、つぎに ATFT の 10 が死滅する。TFT の 01 は、少数になるが生き残り、裏切り型 00 と併存する。

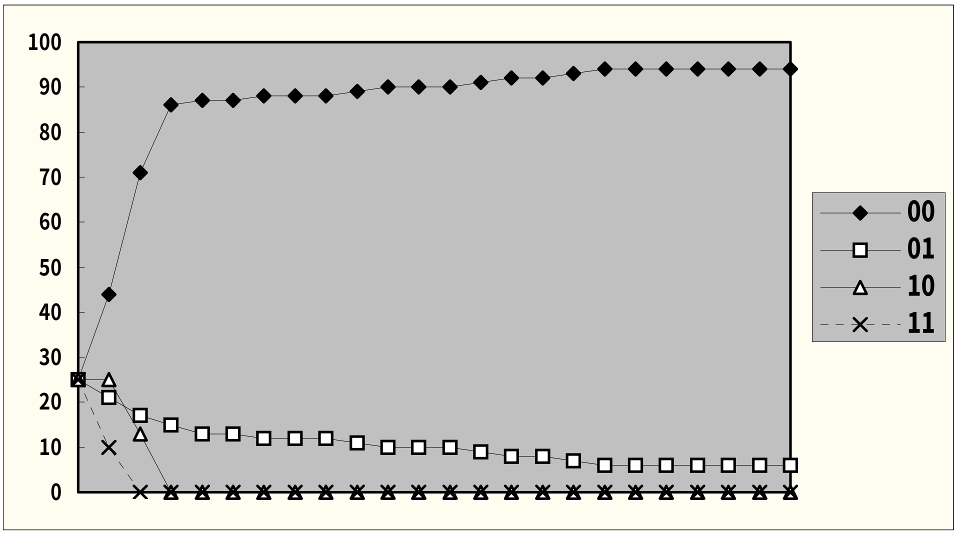

図 1

・ 00 が 10、01 が 80、10 が 5、11 が 5 から開始すると、00 と 01 はやがて死滅する。01 と 11 という正直者だけが生き残ることになる。

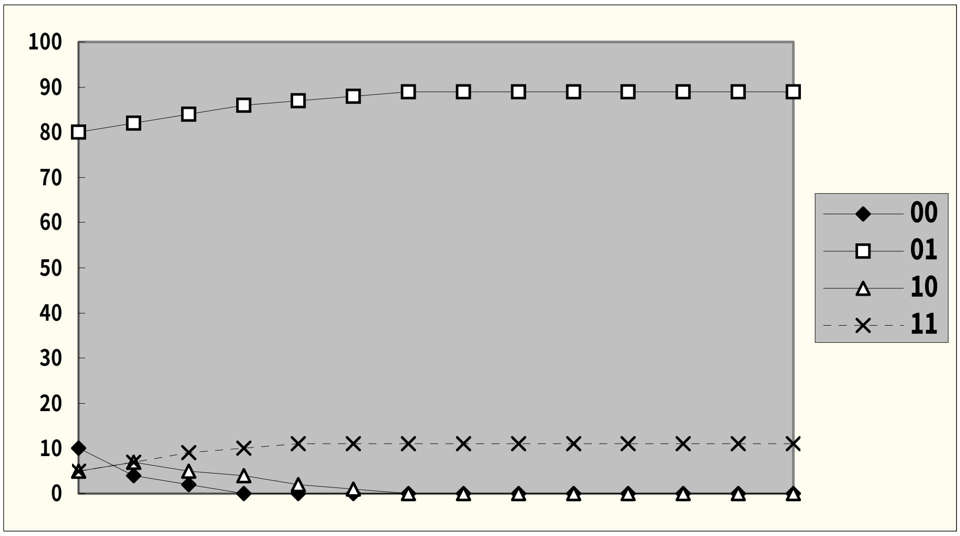

図 2

・ 00 が 15、01 が 60、10 が 15、11 が 10 からスタートすると、はじめは、00 が減少し、01 が増加す

るが、やがて 00 は増加に転じ、11 は死滅し、はじめは増加していた 10 も減少し死滅する。

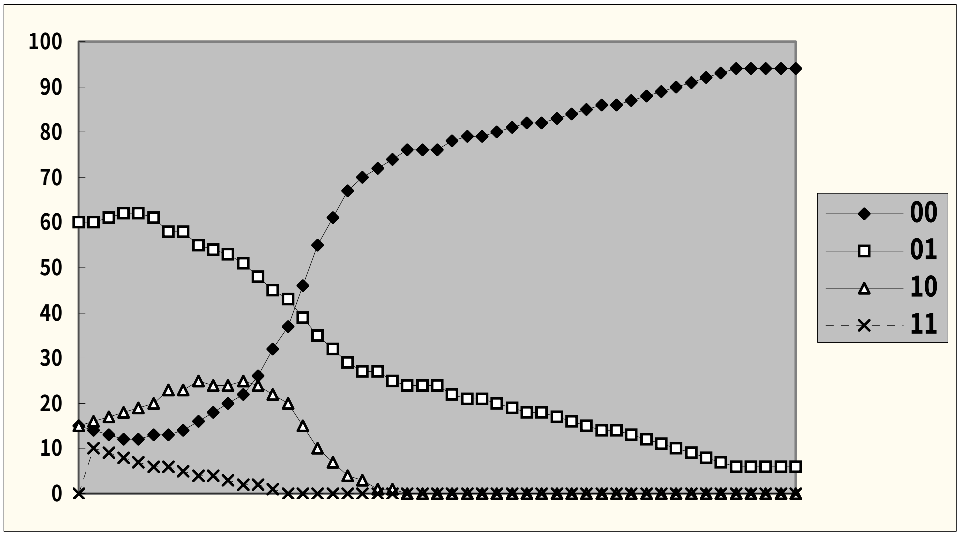

図 3

これら三つの結果から、裏切り戦略（ALLD）が勝つのは、初期にある数だけの正直者（ALLC）や ATFT が存在するときであることがわかる。正直者の数が少なすぎる場合には、裏切り戦略は生き残ることができない。TFT は、裏切り戦略とも正直者とも共存することができる。裏切り者と共存するときは、どちらも裏切り続け、正直者と共存するときは、どちらも協調を選択する。このケースでは ATFT は、どのような初期の個体数から出発しても滅びた。

6. 2 シミュレーション 2（利得行列を変化させた場合）

つぎに利得行列の数値を表 4 のように変更してやる。これは表 6 における裏切ったときの利得 5 と協調したときの利得 3 をそれぞれ 6 と 5 に変更し、裏切ったときと協調したときの利得の差をより小さくしたものである。その結果として、協調する戦略がより生き残りやすくなると考えられる。

プレイヤー 2

プレイヤー 1

	協調（*C*）	**裏 切 り**（*D*）
協調（*C*）	(5,　5)	(0,　6)
裏 切 り（*D*）	(6,　0)	(1,　1)

表 6　囚人のディレンマの利得行列を変化させたもの

・00 を 25、01 を 45、10 を 15、11 を 15 としたとき。このとき TFT(01)の個体数はほとんど変化しな

い。正直者(11)は、その個体数を増やしていき、60 程度で安定する。裏切り者(ALLD)と ATFT はすぐに滅亡する。

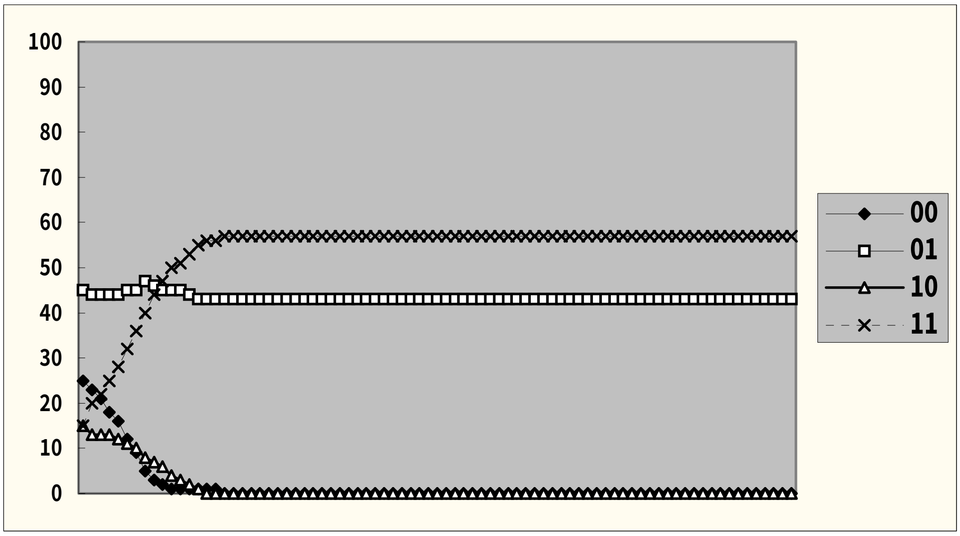

図 4

・00 を 20、01 を 40、10 を 20、11 を 20 としたとき。TFT は 40 前後でほぼ安定して個体数を維持している。ALLD は一時増加するが、ALLC が滅亡することによって再び減少する。ATFT は 20 前後で安定している。

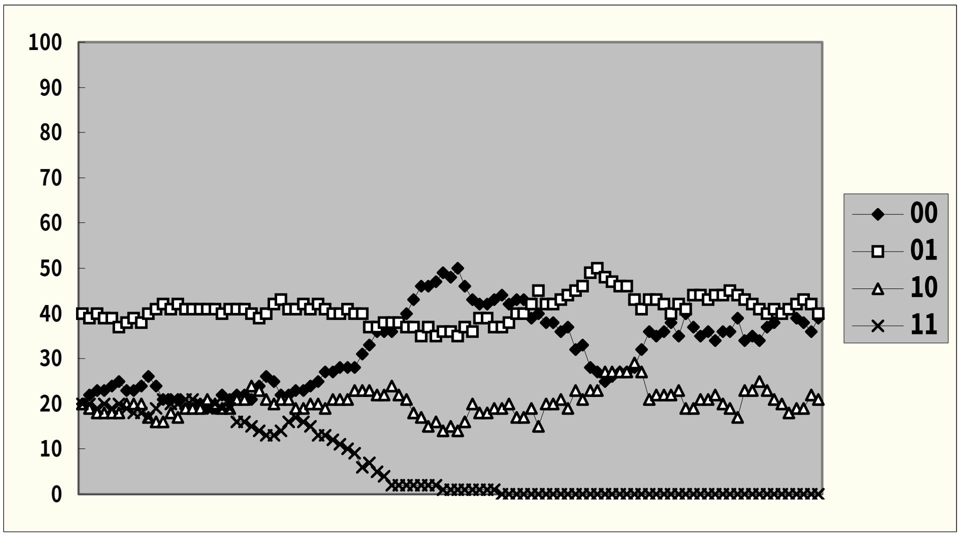

図 5

この結果は、裏切るときに得られる一時的な利益が、協調を取りつづけるときと比較してあまり大きくないときには、裏切り者(ALLD)が弱くなることを示している。TFT は、どのようなケースにおいても強く滅びることはなかった。正直者(ALLC)は、ALLD が存在するときには生き残ることはできない。ALLC が生き残れるのは、TFT との共存のケースだけである。ATFT は、不安定な戦略で、ほとんどのケースについて滅亡した。生き残ったのは、正直者(ALLC)と TFT が共存したケースだけであった。

7. おわりに：今後の発展の可能性

この論文では、メモリーが1で、ノイズが存在しない、また戦略が進化しないという非常に限定されたケースについて、初期の個体数や利得行列の大きさが、戦略の生き残りにどのような影響を与えるのかを分析した。したがって、今後は、分析をノイズが存在するケース、メモリーのサイズがより大きいケース、戦略が進化するケースへと発展させていく必要があるだろう。

まずノイズ、すなわち戦略を誤って行使する可能性が存在するケースについては、どのようなことが新たに生じるのだろうか。TFT と TFT が対戦しているとする。ノイズが存在しないときには、どちらもずっと協調を採用している。しかし、ここで、一方の TFT が協調Ｃしているときに、誤って裏切り D をとるとする。そのときには、以後 D と C が繰り返される。

TFT　CCC　CDCDC

TFT　CCC　DCDCD

つぎに本来 C をとるべきときに誤ると、以後 D が続く。

TFT　CDC　DDDDD

TFT　DCD　DDDDD

さらに、間違うと

TFT　DDD　DCDCD

TFT　DDD　CDCDC

再び裏切り会いになる。一方、本来 D をとるべきときに誤ると、以後 C が続く。

TFT　DCD　CCCCC

TFT　CDC　CCCCC

このようなことを考慮していくと、結局、ノイズが存在するときの利得の期待値は$(R+T+S+P)/4$になる。したがって、TFT 戦略の優位性は、小さいものになることが予想される。このようなケースでは、「相手が続けて 2 回裏切ったときに裏切り返す」という戦略が、ノイズに対しては安定になり、より望ましいのかもしれない。しかし、このような戦略も、1 回おきに協調する相手には負けることになる。また「前回獲得した利得が、R より少なければ行動を変え、そうでなければ行動を変えない」という Simpleton と呼ばれる戦略が強いことが知られている。

つぎにメモリーのサイズが増える場合を考えてみよう。Axelrod が明らかにしたように、メモリーのサイズを増やして、それに戦略の突然変異、交叉など、遺伝的アルゴリズムを組み合わせることによって、より複雑で強力な戦略を産み出すことができる。そのときに、どのような戦略が生き残っていくのかは興味深いテーマである。

参考文献

Axelrod, R. and W. Hamilton (1981), “The Evolution of Cooperation”, *Science* 211, pp. 1390-1396.

Axelrod, R. (1984), *The Evolution of Cooperation*, Basic Books. （松田裕之訳『つきあい方の科学：バクテリアから国際関係まで』HBJ 出版局）

Axelrod, R. (1987). “The Evolution of Strategies in the Iterated Prisoner’s Dilemma”, in L. Davis ed., *Genetic Algorithms and Simulated Annealing*, Morgan Kaufmann.

Binmore, K. (1992). *Fun and Games: A Text on Game Theory*, D. C. Heath and Company.

Fudenberg, D. and J. Tirole (1992). *Game Theory*, The MIT Press.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.

伊庭斉志 (1994)『遺伝的アルゴリズムの基礎：GA の謎を解く』オーム社。

Kreps, D. M. (1990). *Game Theory and Economic Modeling*, Clarendon Press Oxford.

Langton, C. G. ed. (1989). *Artificial Life*, Addison-Wesley Publishing Company.

Langton, C. G. ed. (1993). *Artificial Life III*, Addison-Wesley Publishing Company.

Langton, C. G. ed. (1996). *Artificial Life: An Overview*, The MIT Press.

Maynard Smith, J. (1982), *Evolution and the Theory of Games*, Cambridge University Press. （寺本英・梯正之訳 (1985)『進化とゲーム理論：闘争の論理』産業図書）

Maynard Smith, J. and G. A. Price (1973), “The Logic of Animal Conflict”, *Nature*, 246, pp.15-18.

Poundstone, W. (1992), *Prisoner’s Dilemma*, Doubleday. （松浦俊輔他訳 (1995)『囚人のディレンマ』青土社）

Lindgren, K. (1991), “Evolutionary Phenomena in Simple Dynamics”, in C. G. Langton, J. D. Farmer, S. Rasumussen and C. Taylor eds., *Artificial Life II*, Addison Wesley.

Lindgren, K. and M. G. Nordahl (1993), “Artificial Food Webs”, in Langton(1993).

Lindgren, K. and M. G. Nordahl (1996), “Cooperastion and Community Structure in Artificial Ecosystems”, in Langton(1996).