# Axelrodの繰り返し囚人のジレンマゲーム再考

塩村　　尊*

**要　旨**

Axelrodが行った繰り返し囚人のジレンマゲームに関する一連の数値実験の結果によれば，TFT戦略を現実世界の囚人のジレンマ状況において採用することが推奨される．しかしながら，後の研究により彼のシミュレーション結果は様々な条件に依存しており，TFT戦略の平均的優位性は決して自明のものではないことが明らかになっている．本稿の目的はレプリケータ・ダイナミクスを模した数値実験により，この点を強調するとともに，従来のシミュレーション分析においては，どちらかといえば重要視されていなかったGRIM戦略の平均的優位性とエンフォーサーとしての役割を強調することにある．合わせて，多くのシミュレーション分析において観察される全面協力と全面裏切りの振動を抑える鍵が，上品であるがゆえに高い平均利得を獲得できるエンフォーサーの急成長にあることを示唆する．

キーワード：繰り返し囚人のジレンマゲーム，レプリケータ・ダイナミクス，シミュレーション

# Reconsideration of Axelrod’s Iterated Prisoner’s Dilemma Game

Takashi SHIOMURA

**Abstract**

According to the results of a series of numerical experiments on Axelrod's iterated Prisoner's Dilemma game, people should adopt the tit-for-tat strategy for resolving dilemmas in the actual world. The success of this strategy is, however, by no means obvious, since the results depend heavily on various conditions of the experiments. This study reconsiders Axelrod's computer tournament and ecological system, as well as reconfirms that his results depend, in particular, on the set of strategies played in the game and their initial distribution. Moreover, we emphasize the success of the grim trigger strategy and its importance as an enforcer that suppresses the prosperity of naive cooperators and cunning defectors. The key to damping an oscillation, which appears in many simulations (and generates perpetual alternation between mutual cooperation and mutual defection), may be the rapid proliferation of *nice* enforcers which are to say never being the first to defect.

Key words: Iterated prisoner's dilemma game, replicator dynamics, simulation

* 関西大学総合情報学部

## 1．Axelrod の繰り返し囚人のジレンマゲーム

相手プレイヤーとの無限回の対戦を念頭に置き，利得の割引現在価値の最大化という観点から行動計画，すなわち戦略を決定するゲームを繰り返しゲームと呼ぶが，とくに成分ゲームとして囚人のジレンマゲームが与えられている場合の繰り返しゲームを繰り返し囚人のジレンマゲーム（以下，IPD ゲーム）と呼ぶ．Axelrod[1]-[3]が行った IPD ゲームに関する数値実験と代数解析によれば，1)TFT 戦略に代表される上品，かつ毅然と寛容のバランスのとれた，相手に態度が分かり易い戦略が平均としてよいパフォーマンスをあげること，および 2) 割引因子が十分大きいという条件の下，TFT 戦略は集団安定，かつ領域安定な戦略になることが確認されており，これらの事実から彼は結果的に 2 人のプレイヤーが互いに協調戦略をとり続ける TFT 戦略を現実世界の囚人のジレンマ状況においても採用することを推奨している．

しかしながら，後の研究により Axelrod による IPD ゲームに関する一連のシミュレーション結果は様々な条件に依存しており，TFT 戦略の平均的優位性は決して自明のものではないことが明らかになっている．本稿の目的は Linster[4]によるレプリケータ・ダイナミクスを模した数値実験のいくつかを再現をして Axelrod の結果と比較し，この点を強調するとともに，従来のシミュレーション分析においては，どちらかといえば重要視されていなかった GRIM（TRIGGER）戦略の平均的優位性とエンフォーサー（enforcer）としての役割を強調することにある．合わせて，多くのシミュレーション分析において観察される全面協力と全面裏切りの振動を抑える鍵が，上品であるがゆえに高い平均利得を獲得できるエンフォーサーの急成長にあることを示唆する．

## 2．総当たり戦

以下の分析において考える戦略は遺伝子型で表現したときに遺伝子長 7 で表すことができる 32 戦略のうち，表現型の異なる 26 戦略である．詳細については補足を参照されたい．各遺伝子座の意味は表 1 のとおりである．以後は 2 × 2 成分ゲームにおける各プレイヤーの戦略，すなわち協調と裏切りを各々，文字 C と D，または 2 値 0 と 1 で表す．

表 1　遺伝子座の意味

| 遺伝子座 | 意味 |
|---|---|
| 0 | 前回プレイの自身の手 |
| 1 | 前回プレイの相手の手 |
| 2 | 初回の自身の手 |
| 3 | 記憶部が（C, C）=（0, 0）の場合の自身の手 |
| 4 | 記憶部が（C, D）=（0, 1）の場合の自身の手 |
| 5 | 記憶部が（D, C）=（1, 0）の場合の自身の手 |
| 6 | 記憶部が（D, D）=（1, 1）の場合の自身の手 |

成分ゲームの利得表は表2のとおりであり，これはAxelrod[1],[2]が用いたものと同じである．各戦略は自身を含めた戦略すべてと同数回対戦するものとし，そのときに獲得できる利得の割引現在価値の平均値によって戦略を評価する．なお，割引因子の大きさは本稿を通じて0.99で固定されている．

表2　成分ゲームの利得表

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

図1は以下の議論において重要な役割を果たす7つの戦略の平均利得の推移を観察したものである．これら7戦略の意味は表3のとおりである．図から明らかなように平均利得は自身も含めたすべての戦略との対戦回数に依存するが，とくにALL-D戦略の平均利得の推移に注意されたい．この戦略は対戦回数が少ない場合には最高の平均利得をあげることができるが，十分な対戦回数を念頭に置いた場合は，換言すれば十分先の対戦結果を考慮にいれた場合は26戦略中，最終的に第1位と2位の成績を収めるGRIMやSOURPUSS戦略の平均利得を下回り，第3位になる．Axelrod[1], [2]においては優秀な成績を収めたTFT戦略の平均利得が低いことにも注意されたい．実際，TFT戦略は26戦略中の9位にすぎない．第2位と3位の戦略が，自ら先んじて裏切ることがない，いわゆる「上品な」戦略ではないことも注目に値する．

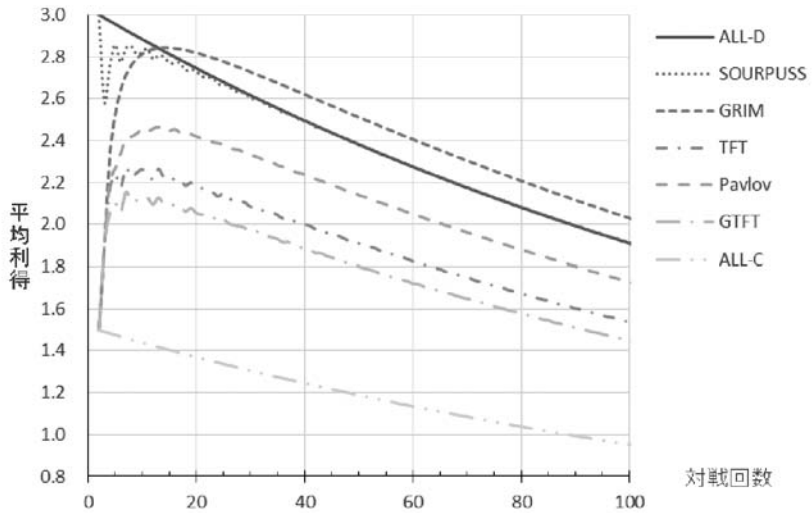


図1　対戦回数と平均利得

表3　7戦略の意味：戦略番号は補足を参照されたい.

| 名前 | 戦略番号 | 意味 |
|---|---|---|
| ALL-D | 1 | 常に裏切る. |
| SOURPUSS | 9 | 初回はDであり，前回プレイで少なくとも1人がCをとるときにはDを返し，両者がDのときのみCを返すひねくれ者（sourpuss）の戦略. |
| GRIM | 18 | 相手がD戦略をとらない限りCを返すが，一度裏切られると永久にDを返す容赦のない（grim）の戦略. |
| TFT | 22 | 初回はCで，2回目以降は相手の戦略を採用する. |
| Pavlov | 26 | 初回はCで，2回目以降は前回プレイの結果がともに協調か裏切りならばCを返し，そうでなければDを返す. |
| GTFT | 30 | 初回はCで，2回目以降は前回プレイの結果が，自分がCで相手がDであるときのみDを返す寛容な（generous）TFT戦略. |
| ALL-C | 32 | 常に協調する. |

詳細は省略するが，補足にあげた遺伝子型の異なる32戦略すべてで総当たり戦を行った場合はALL-D戦略が最高の平均利得を獲得するようになる．この理由は32戦略中，4つの戦略が実質的にALL-C戦略になるために，ALL-D戦略がこれら「お人好し」を食い物にし，高い利得を獲得することができるためである．他方，26戦略からGRIM戦略を除いた25戦略で総当たり戦を行うとSOURPUSS戦略が最高得点を獲得するようになる．これらの事実は，いかにAxelrodの総当たり戦の結果が考察する戦略の性質と初期分布に強く依存していたかを示唆している.

## 3．生態学的模擬実験

つづいて，レプリケータ・ダイナミクスを模した実験を行う．これは以下のように実装されている．今，$N_i^t$と$W_i^t$を各々，$t$世代，$t \geq 0$において第$i$戦略，$i = 1,2,...,n$をとる人数と利得を表すものとする．このとき，$t+1$世代において第$i$戦略をとる人数，$N_i^{t+1}$は

$$N_i^{t+1} = N_i^t \left( \frac{W_i^t}{\overline{W}^t} \right) \tag{1}$$

に従って変化するものと仮定する．ここで，

$$\overline{W}^t \equiv \sum_i N_i^t W_i^t \Big/ \sum_i N_i^t$$

であり，これは$t$世代におけるプレイヤーの平均利得を表している．また，単純化のために1人の$i$プレイヤーが1人の$j$プレイヤーと対戦したときに獲得できる利得，$S_{ij}$は世代を通じて同じであるとし，

$$W_i^t \equiv \sum_j S_{ij} N_j^t$$

により定義する.

さて，列和が1の確率行列$M \equiv (m_{ij})$を定義する．ただし，$i, j = 1,2,...,n$である．このとき，

突然変異確率を $0 \leq \mu \leq 1$ として

$$m_{ij} \equiv \begin{cases} 1-(n-1)\mu, & i=j \\ \mu, & i \neq j \end{cases}$$

を定義する．もちろん，$0 \leq m_{ij} \leq 1$ であるように $\mu$ が選ばれなければならない．我々は次のようにして突然変異を生み出している．すなわち，$t+1$ 世代の戦略人口が式(1)によって確定した後，

$$N_i^{*t+1} = \sum_j m_{ij} N_j^{t+1} \tag{2}$$

によって $t+1$ 世代における突然変異後の戦略人口，$N_i^{*t+1}$ を決定している．式(2)は文献[4]，p. 889に従ったものであるが，世代ごとに複数の突然変異が同時に現れることに注意されたい．また，26戦略と異なる，まったく新しい戦略を生み出すわけではないことにも注意されたい．この意味において，突然変異と呼ぶよりは再生と呼ぶ方が適切かもしれない．

### 3.1 突然変異が存在しない場合

各戦略は自身も含めてすべての戦略と100回対戦するものとして，それぞれの戦略が獲得できる利得の割引現在価値の1回当たり平均値を $S_{ij}$ と定義する．図2は26戦略の初期戦略人口をそれぞれ100としたときの力学系(1)が生成する100世代後における戦略分布であるが，実は約20世代後には平衡状態に達する．最終的に生き残った5つの戦略はすべて上品な戦略であり，高い平均利得をあげることができるが上品ではない2つの戦略，すなわちSOURPUSSやALL-Dは淘汰圧の中で生き残ることができない．また，平均利得が26戦略中25位であったALL-C戦略が，わずかながらも生き残っていることは注目に値する．これらの事実は文献[4]，p. 892,

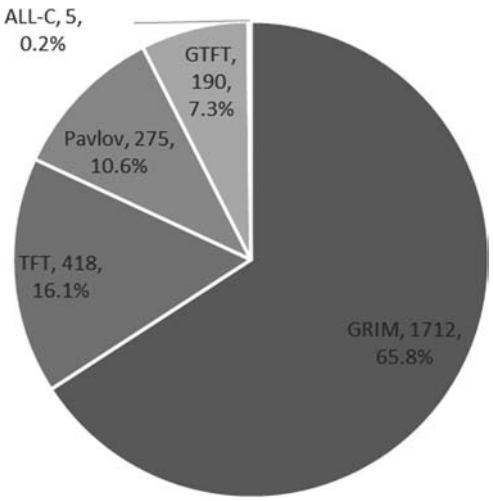


図2　最終的な戦略分布：突然変異無しの場合

Figure 7の結果とほぼ一致する．なお，Linster[4]の戦略ccとcdは各々，我々のPavlov，GTFTと同じものである[1)]．

一方，図3は20世代までの上記7つの戦略人口の推移を表したものである．図より初期段階においては高い平均利得を反映して分布を拡大していたALL-DとSOURPUSS戦略がGRIM戦略の急成長とともに，勢力を失っていくことが確認できる．平衡状態において，これら2つの上品ではない戦略が消滅してしまった理由は，GRIM戦略が持つエンフォーサー（相手の裏切りを発見し，これを罰する者）としての性質がこれら2つの戦略の繁栄を許さないためであったと考えられる．また，これら2つが消滅してしまい，集団が上品な戦略のみになってしまった中では，お人好しのALL-C戦略も細々と生き残ることができるのである[2)]．

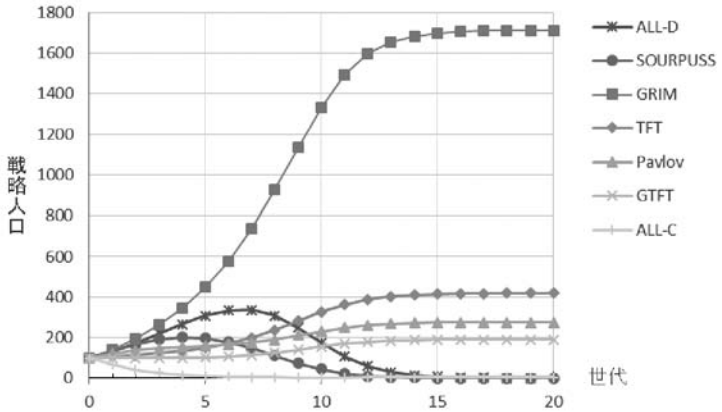


図3　戦略人口の推移：突然変異無しの場合

以上の考察から，初期状態において前回プレイのみに依存する26戦略が均一に分布し，かつ突然変異が存在しないという前提の下，最終的には複数戦略の共存と相互協力が実現することがわかる．

### 3.2　突然変異が存在する場合

図4は式(2)における突然変異確率，$\mu$を0.0004に設定したときの100世代後における戦略分布であるが，やはり約20世代後にはほぼ平衡状態に達する．ただし，突然変異を反映して若干の変動は起こり続ける．

---

1）文献[4]，p. 891，Figure 6には誤りがある．図中のcbとccはそれぞれ，TFTとPavlovに対応する．また，GRIMに対応するのはcaである．

2）GRIM戦略を除く25戦略で実験を行った場合はTFT，Pavlov，およびGTFTの3つの戦略が生き残り，TFT戦略が集団中の大勢を占めるようになる．

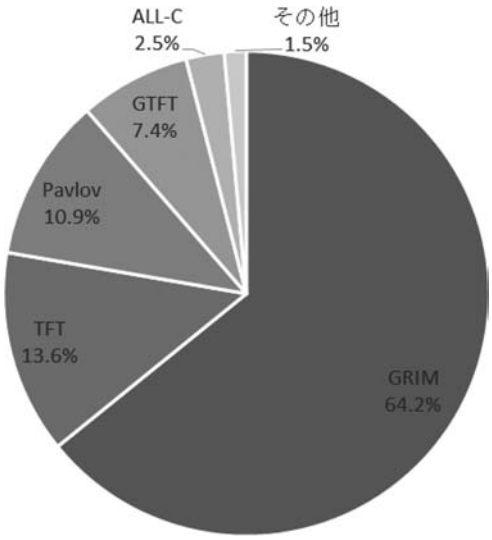


図 4　最終的な戦略分布：突然変異有りの場合

図より突然変異を導入したことにより，突然変異が無い場合よりも ALL-C 戦略が多く生き残っていることが確認される．この理由は我々が定式化した突然変異，すなわち式(2)より，少数ながらも ALL-C 戦略とうまくつきあうことのできる多くの上品な戦略を新たに生み出すと同時に，同じく新たに生み出される上品ではない戦略の繁栄を GRIM 戦略等のエンフォーサーが抑えるためであると考えられる．この結果は文献[4]，p. 893，Figure 8 と傾向としてほぼ一致している．

ところで，Axelrod 以降の進化論的アプローチをとる数値実験ではしばしば全面協力と全面裏切りの永続的振動が起こることが報告されている（たとえば，文献[5]–[7]参照）．一方で，最大で10万世代におよぶ我々の数値実験ではこのような振動は観察されなかった．この理由は Linster[4] と我々のモデルにおいては，割引因子が十分に大きいという前提の下，1)平均利得が低いがゆえに ALL-D 戦略を集団中に招き入れ，全面協力から全面裏切りへの転換のきっかけを与える ALL-C 戦略等のお人好しが比較的早い段階で衰退すること，および2)平均利得の高い GRIM 戦略が集団内で急成長するために ALL-D 戦略等の上品ではない戦略の繁栄が許されなかったためであると考えることができる．さらにいえば，我々と Linster が用いた突然変異のメカニズム，すなわち式(2)が実は確率的なものではないことも振動が起こらない 1 因になっているように思われる．実際，Linster の数値実験において振動が起こるのは ALL-C や ALL-D 戦略が成長するような式(2)とは異なるメカニズム，すなわち文献[8]，pp. 149–152 による戦略の複雑さに関するコストを考慮したメカニズムがモデルに組み込まれている場合である（文献[4]，pp. 896–899 参照）．戦略の成長メカニズムがどのようなものであれ，全面協力と全面裏切りの振動を抑える鍵はお人好しと裏切り者の繁栄を許さない，上品なエンフォーサーの急成長であ

るように思われる．

## 4．結語

以上，我々はレプリケータ・ダイナミクスを模した数値実験を行い，Axelrodが得た一連の結果が戦略の性質と初期分布に強く依存していることを再確認してきた．また，前回プレイのみに依存する純粋戦略を考える限りにおいて，GRIM戦略が最高の平均利得を獲得するために集団中の多数派となることが確認された．Linster[4]の様々な数値実験においてGRIM戦略が優勢になる決定的理由はこれである．ただし，この結果もまたAxelrodの数値実験と同様に戦略の性質と初期分布に強く依存していることを強調しておく．

Axelrod以降の進化論的観点に立つ数値実験ではしばしば全面協力と全面裏切りの永続的振動が起こることが報告されているが，我々の数値実験ではこのような振動は観察されなかった．この理由は裏切り者の侵入を許してしまう，お人好しが比較的早い段階で衰退する一方で，急成長するGRIM戦略が裏切り者の繁栄を抑えたためであると考えることができる．このことは全面協力と全面裏切りの振動を抑える鍵が，上品であるがゆえに高い平均利得を獲得できるエンフォーサーの急成長にあることを示唆しているように思われる．

## 補足

表Aは我々が用いた戦略の一覧であり，文献[7]，p. 105，Table 2と実質的に同じものであるが，そこでは表Aにおける7と10の戦略が欠落している．また，同Table 2の戦略6と9の説明は各々，これら欠落した7と10に関するものである．表中，戦略2から4までは戦略1（ALL-D）と，戦略20，24，および28は戦略32（ALL-C）と遺伝子型は異なるものの，表現型が同じであることに注意されたい．それゆえ本稿におけるプログラムの実装では，これら6戦略が除かれている．

表A　戦略一覧：1から16までが初回がDであり，17から32までが初回がCである．

| No. | 遺伝子型 | 2回目以降の戦略 |
|---|---|---|
| 1 | 11111 | 常にD |
| 2 | 10111 | 常にD |
| 3 | 11011 | 常にD |
| 4 | 10011 | 常にD |
| 5 | 11101 | (D, C) のときのみC |
| 6 | 10101 | TFT |
| 7 | 11001 | 逆Pavlov |
| 8 | 10001 | 逆GRIM |
| 9 | 11110 | (D, D) のときのみC |
| 10 | 10110 | Pavlov |
| 11 | 11010 | 相手と反対の戦略 |
| 12 | 10010 | (D, C) のときのみD |
| 13 | 11100 | 自分の手を交互に変える |
| 14 | 10100 | (C, D) のときのみD |
| 15 | 11000 | (C, C) のときのみD |
| 16 | 10000 | 常にC |
| 17 | 01111 | 常にD |
| 18 | 00111 | GRIM |
| 19 | 01011 | (C, D) のときのみC |
| 20 | 00011 | 常にC |
| 21 | 01101 | (D, C) のときのみC |
| 22 | 00101 | TFT |
| 23 | 01001 | 逆Pavlov |
| 24 | 00001 | 常にC |
| 25 | 01110 | 2回目以降は (D, D) のときのみC |
| 26 | 00110 | Pavlov |
| 27 | 01010 | 相手と反対の戦略 |
| 28 | 00010 | 常にC |
| 29 | 01100 | 自分の手を交互に変える |
| 30 | 00100 | (C, D) のときのみD |
| 31 | 01000 | (C, C) のときのみD |
| 32 | 00000 | 常にC |

## 参考文献


［1］ Axelrod, R.: Effective choice of Prisoner's Dilemma, *Journal of Conflict Resolution*, Vol.24, No.1, pp.3–25 (1980).

［2］ Axelrod, R.: More effective choice of Prisoner's Dilemma, *Journal of Conflict Resolution*, Vol.24, No.3, pp.379–403 (1980).

［3］ Axelrod, R.: The emergence of cooperation among egoists, *American Political Science Review*, Vol.75, pp.306–318 (1981).

［4］ Linster, B. G.: Evolutionary stability in the infinitely repeated Prisoners' Dilemma played by two-state Moore machines, *Southern Economic Journal*, Vol.58, pp.880–903 (2001).

［5］ Lindgren, K.: Evolutionary phenomena in simple dynamics, in C. G. Langton, J. D. Farmer, S. Rasmussen and C. Taylor eds.: *Artificial Life II*, pp.295–312, Addison-Wesley, Redwood City CA (1991).

［6］ Nowak, M. A. and Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's

Dilemma game, *Nature*, Vol.364, pp.56–58 (1993).
［7］ Hoffmann, R.: The ecology of cooperation, *Theory and Decision*, Vol.50, pp.101–118 (2001).
［8］ Rubinstein, A.: *Modeling Bounded Rationality*, The MIT Press, Cambridge MA (1998).