

リーダーシップの形成に関する進化ゲーム — シミュレーションアプローチ

FORMATION OF LEADERSHIP — SIMULATION APPROACH

秋山英三¹
Eizo Akiyama

筑波大学大学院システム情報工学研究科京都大学大学院 情報学研究科, 〒 305-8573 茨城県つくば市天王台 1-1-1
Graduate School of Systems and Information Engineering, University of Tsukuba

概要

エラー付きの「繰り返し指導者ゲーム」に関する進化シミュレーションを行った。エラーなしのケースに関する既存研究の結果と異なり、指導者と追従者の役割を固定化する方向の進化の過程が見られた。また、生き残った戦略にはすべてのシミュレーション試行について共通のポリシーが見られた。それは次の三つのポリシーである。(1)「上手く行動をコーディネートできない相手に、自分からは譲らない」(2)「いったん指導者の経験をする」と積極的に振る舞う (3) 自分が指導者になった記憶がない場合、相手が常に積極的に指導者の経験もあるなら相手に追従する。

1 指導者ゲーム

目的地へと急いでいる二台の自動車が、交差点で渋滞の流れに割り込もうとしている状況を考えてみよう。それぞれの運転手は、車の流れに少し隙間ができた時に、積極的かつ大胆に隙間に飛び込む (D=Dive, Daring) か、慎重に相手の車に譲るか (C=Concede, Careful) を決断する。もし二台とも譲ってしまうと二台とも遅刻してしまう (2 番目に悪い結果)。もし二台ともあわてて飛び込むと事故が起こる (最悪の結果)。片方が隙間に飛び込んで、もう片方がそれに続けば、飛び込んだ方 (リーダー) が早く目的地に到着し、ほんの少し遅れてもう一台が到着する。このようなゲームのことを「指導者ゲーム [5]」と呼ぶ。このゲームを利得行列で表現すると、例えば表 1 のようになる。

表 1 指導者ゲーム (The Game of Leader) の利得行列: それぞれのセルにおいて、左の数値は行プレイヤー (プレイヤー 1) の利得、右の数値は列プレイヤー (プレイヤー 2) の表す。

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

このゲームには、C (消極的)、D (積極的) という二つの選択肢があり、二人のプレイヤーが別々の選択肢を選んだ状態が純粋戦略ナッシュ均衡でありパレート最適でもある。このような役割分業が成立したとき、本稿では D を選択している方を「指導者」、C を選択している方を「追従者」と呼ぶことにする。このゲームで最も得るのは指導者になることだが、そう思って両プレイヤーが積極的に行動してしまうと最悪の結果に陥ることにな

る¹。

指導者ゲームは Rapoport が分類した non-trivial な二人二戦略 (2x2) ゲームのうちの一つであり、その他のゲームの中には囚人ジレンマなどが含まれる [5]。繰り返し 2x2 ゲームにおける進化については、例えば、Crowley [2] が、前回の相手と自分の行動の記憶を用いる記憶長 2 の classifier system を用いて 1000 世代の進化を分析している。また Tanimoto ら [6] は、2x2 ゲーム全体をパラメータ表現することに成功し、ゲームの構造が、記憶長 2 の戦略の進化とゲームでの振る舞いに与える影響を総括的に検証している。(現実の指導者ゲーム状況の例はこれらの論文の中でも紹介されている)

表 1 では、(C,C) の時の利得 3 点の 2 倍、つまり 6 点よりも、(C,D)(D,C) の時の利得の和、つまり 3+5=8 点の方が大きいことに注意しよう。Browning ら [1] はこのようなゲームにおいて、二人が (C,D) 状態と (D,C) 状態を毎回交代する Alternating reciprocity が進化することを記憶長 6 の遺伝的アルゴリズムによって示した²。

一方、現実世界のプレーヤーにミス (エラー) は避けられないので、例えば、囚人ジレンマについては、エラー付き繰り返し囚人ジレンマの理論あるいはシミュレーション研究が数多く行われている (例えば [3],[4],[7])。

では、エラー付きの繰り返し指導者ゲームにおける良い戦略とはどのような戦略だろうか? 本稿では、エラー付き繰り返し囚人ジレンマを分析した Lindgren [3] のモデルをベースとして、エラー付きの繰り返し指導者ゲームにおける「良い」戦略を探る。次節で述べるように、各試行では 10 万世代の進化シミュレーションを行う。突然変異によって記憶長が変化するようにし、最大記憶長は 8 とした。分析するゲームは指導者ゲームに限っているが、途中の戦略進化の過程・論理を詳細に検証する。

¹指導者ゲームの状況は、(1) 行動のコーディネーションがグループ内で行われると皆にメリットがあるが、(2) そのメリットは平等ではなく、積極的な者が得をするような状況で現れる。例として、捕食者から逃げる 2 匹の生物が、一度に一匹しか通ることができない逃げ道に行き着いた状況を考えよう。この場合、2 匹が行動を調整して、片方が先に通路を通り、もう片方がその後を追いかける状態が望ましい。最悪なのは、両方が通路に飛び込んで傷つけ合い、捕食者の餌になることである [1]。また、多くのグループワークでは、グループにリーダーがいる方が生産性が上がる。この時、リーダーになることに特権的な利益があると指導者ゲーム状況になる。(リーダーにぶら下がる人の方にメリットが大きい状況は「英雄ゲーム」と呼ばれる。)

²[1] の設定では、151 ラウンドの繰り返し 2x2 ゲームにおける進化を、population サイズが 20 の総当たり戦を前提とした 1000 世代の遺伝的アルゴリズムによって分析している。

2 モデル

本稿で用いるモデルにおけるシミュレーションの手続きとパラメータの値は、K. Lindgren (1992) の繰り返し囚人ジレンマゲーム進化シミュレーションの方法 [3] に基本的に基づいている。

2.1 繰り返し指導者ゲーム

プレーヤーたちは指導者ゲーム (表 1) を T ラウンド繰り返し行なう。(つまり、 T =最大ラウンド数。) 以下、繰り返しゲームの一回分のゲームのことを以下「ステージゲーム」と呼び、各ステージゲームにおけるプレーヤーの選択肢のことを「行動」と呼ぶことにする。各プレーヤーは有限記憶長の決定論的「戦略」を持ち、過去のステージゲームにおける両プレーヤーの行動の履歴から次のステージゲームでの行動を決定する。つまり、プレーヤーの戦略は、過去の有限長の「履歴」を状態として次の行動を決定するムーアマシンである。

なお、本稿では、プレーヤーの行動にエラーが起こることが仮定される。つまり、プレーヤーは、一定の確率 p で、意図したのと逆の行動を選択してしまう (間違える)。また、プレーヤーが間違える時間間隔の期待値は、最大ラウンド数より十分に小さく、記憶長 m より十分大きいことが仮定される ($T \gg 1/(1 - (1 - p))^2 \gg m$)。つまり、エラーはたまにしか発生しないが、 T ラウンドが終わるまでの間には十分な回数のエラーが発生する³。

各プレーヤーが記憶する記憶長 m の履歴は、対戦相手の前回の行動 a_0 、自分の前回の行動 a_1 、対戦相手の前々回の行動 a_2 、自分の前々回の行動 a_3 、... のように両プレーヤーの過去の行動の列で構成される。行動 D を 0、行動 C を 1 と書くことで、長さ m の履歴 h_m を $h_m = (a_{m-1}, \dots, a_1, a_0)_2$ のように二進コードで表現することができる。

記憶長 m のプレーヤーが出会う可能性がある履歴は $(0, \dots, 0, 0)_2, (0, \dots, 0, 1)_2, \dots, (1, \dots, 1, 1)_2$ の 2^m 種類ある。本稿では決定論的戦略を仮定しているので、これらの履歴それぞれに出会ったときに行う行動 $A_0, A_1, \dots, A_{m-1}$ は一意に決まる。従って、記憶長 m の戦略は、長さ m の二進コード $S = [A_{m-1}, \dots, A_1, A_0]$ で表現することができる⁴。例えば、記憶長 $m = 1$ の二進コード $[10]$ で表現される戦略は、前回相手が C なら C を、D なら D を選択する戦略である。「記憶長 m の戦略」の種類の総数は 2^m である。以下、この二進コードの戦略のことをこの節ではゲノムと呼ぶことにする。

この繰り返しゲームにおける各プレーヤーのパフォーマンスは T ラウンドの平均利得によって決まる。 $T \gg 1/(2p - p^2) \gg m$ を仮定しているので「1 ラウンド目の

行動の選択」が平均利得に与える影響は微小である。[3] では $T \rightarrow \infty$ を仮定することで 1 ラウンド目の選択の影響を排除し、さらに、有限記憶長履歴の定常分布を考えることで平均期待を解析的に計算していた。本稿の全てのシミュレーションでは、 $T = 10,000, p = 0.01$ と設定し、また、1 ラウンド目の選択をランダムとして実際に繰り返し対戦させ、ラウンド当たりの平均利得を計算している⁵。

2.2 進化ダイナミクス

「繰り返し指導者ゲーム」が行われる世界で、どのような戦略が現れ広がっていくかを見るために、遺伝子 (戦略) の突然変異による変異の創造、遺伝、自然選択のメカニズムを導入し、進化のダイナミクスを分析する⁶。

まず、 N 個体からなる population を考える。(本稿では N は 1000 に設定している。) population 内の個体は他個体と出会うと繰り返し指導者ゲームを行う。

まず、同じ戦略を持つ個体は同じ「種族」に属すると考える。種族 i の個体数は、1.0 を全体にした割合 (頻度) x_i で表される。個体は同族集団も含めたすべての個体と戦う⁷。種族 i, k に属する個体どうしが、繰り返しゲームを行なった結果、1 ラウンドあたりで種族 i の個体が獲得する平均利得を g_{ij} とすると、種族 i の個体が population の全ての個体とゲームを行なって獲得する利得は $s_i = \sum_j g_{ij} x_j$ である。従って、システム全体の平均利得は $s = \sum_i s_i x_i$ となる。平均利得がシステム全体の平均を上回るかどうかで各種族の適応度を測る。種族 i の適応度は $w_i = s_i - s$ で表される。そして、世代 t から $t+1$ にかわるにあたって、適応度に応じて次の世代の各種族の頻度が $x_i(t+1) - x_i(t) = d w_i x_i(t) + m_i$ に従ってきまる。ここで、 d は growth constant で本稿では 0.1 に設定している。また、 m_i は突然変異による頻度の増減の項だが、これについては後述する。

平均利得が population 全体の平均 s より低い種族は次の世代で頻度を減らすことになるが、その時、頻度がしきい値 $1/N$ より少ない種族は絶滅する。絶滅した種族が現れたら全体が 1.0 になるように頻度を正規化する。

突然変異には (1) point mutation (2) gene duplication (3) split mutation の三種類がある。point mutation では、 $[0010] \rightarrow [0110]$ のように、ゲノム (戦略) の中のシンボルが変化する。gene duplication では $[10] \rightarrow [1010]$ のように、自分自身のコピーを追加する。この過程で戦略の内容は変わらないまま記憶長が長くなっていることに注意しよう。split mutation は $[0011] \rightarrow [00]$ のように、ゲノムを前半部と後半部に分割し、どちらかをラ

³ 実行時に意図したのと逆の行動を選択してしまうことを implementation error、相手の行動を間違って知覚してしまうことを perception error 呼ぶことが多い。

⁴ Lindgren の原論文では、左端から $S = [A_0, A_1, \dots, A_{m-1}]$ となっている。本稿では履歴番号を二進コードの桁数にしたかったので逆順に変更したが、分析に本質的な違いは全くない。

⁵ 10,000 ラウンド中に約 200 回のエラーが発生するので初期行動の影響はほとんど無い。実際、囚人ジレンマの利得行列を用いるとシミュレーションの結果が Lindgren の結果とほぼ同様の傾向を持つことを確認している。

⁶ 本稿で用いている進化ダイナミクスは基本的に離散時間レプリケータと呼ばれるものだが、標準的レプリケータと異なるのは、戦略の数が突然変異と淘汰によって増減しうるところである。

⁷ 対戦確率が戦略頻度に比例する well mixed な population なら、ランダムに対戦すると考えても同様の結果になる。

ンダムに捨てる。本稿では、point mutation が発生する確率は各シンボルごとに $p_p = 2.0 \times 10^{-5}$ 、(2)gene duplication と (3)split mutation が発生する確率はそれぞれ $p_g = p_s = 1.0 \times 10^{-5}$ としている⁸⁾。

3 指導者ゲームにおける戦略の進化

以上の設定で、「繰り返し指導者ゲーム」の10万世代の進化シミュレーションを行った。初期世代は、記憶長1の4戦略[00][01][10][11](繰り返し囚人ジレンマではそれぞれ AllD, Anti-TFT, TFT, AllC と呼ばれる戦略)が population の1/4ずつを占める状態から始めた。シミュレーションは、突然変異時の乱数によって試行ごとに異なる結果に分岐していく。gene duplication によって記憶長の伸張が可能だが、計算機の記憶領域の制限から最大記憶長は8に制限した。ただし、現れてすぐに絶滅せずに残ったのは記憶長6までの戦略で、現れた戦略のほとんどは記憶長5までだった。

10万世代までのシミュレーションを8試行行ったが、全体的な傾向は大きく分けて二通りある。一つは図1-(a)のように記憶長が4を超えて4と5の間を振動するもの、もう一つは図1-(b)のように記憶長が4以上にならないものである。前者では、平均利得が3.1点前後と3.9点前後を往復し、後者では3.1点前後に(10万世代の段階で)落ち着いている。

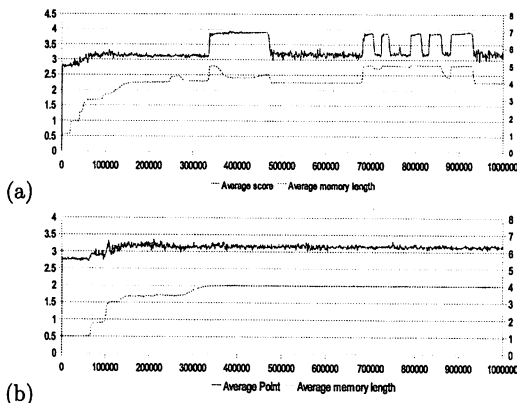


図1 populationにおける平均利得と平均記憶長の遷移の典型例。(a)(b)とも横軸は世代。左側の縦軸は各世代の平均利得、右側の縦軸は平均記憶長の値を表し、前者のデータは実線、後者のデータは点線で示されている。シミュレーションの結果は、大きく分けて、(a)のように記憶長が4を超えて4と5の間を振動するものと、(b)のように記憶長が4以上にならないものの二通りのがある。

⁸⁾なお、ゲノムの各サイトを遺伝子座とみなし、各シンボルを対立遺伝子とみなすと、実際の生物とほぼ同じ突然変異率になる。

⁹⁾以上三つの突然変異によりゲノム j からゲノム i への突然変異が起る確率を q_{ji} とすると、ゲノム j を持つ個体の数は Nx_j なので、ゲノム j が i から j に変化する個体の数の期待値は $Nx_j q_{ji}$ である。ゲノム j からゲノム i を持つ個体が一つ現れる場合に1をとる確率変数を Q_{ij} とすると、本稿のように突然変異率非常に小さく $p_p + p_d + p_s \ll 1/N$ となる場合、 $P(Q_{ij} = 1) = Nx_j q_{ji}$ と近似できる。従って、突然変異による頻度の増減の項 m_i は $m_i = \frac{1}{N} \sum_j (Q_{ij} - Q_{ji})$ となる。

以下では、図1-(a)のシミュレーションで見られた戦略の進化の概要を時間(世代)に沿って紹介し、(1)エラー付き指導者ゲーム世界における戦略進化にはどのようなパスが有り得るのか、(2)(10万世代の時点で)どのような戦略が優勢になるのかを考察する。

(1)に関しては、突然変異の起こり方によって進化の速度・様相が異なり、シミュレーションごとと同じ部分と異なる部分がある(あるシミュレーションのある世代で現れた戦略が他のシミュレーションではスキップされるなど)ので、当然ながら、一つの進化のパスの紹介ということになる。エラーのあるゲームの相互作用を通じてムーアマシンがどのように複雑化していくのかを分析した例としては、エラー付き囚人ジレンマについてはLindgrenの著名な研究があるが、エラー付き指導者ゲームに関しては本研究が初めてである。(2)に関しては全シミュレーションに共通する性質で、おそらく、一定の普遍性があるものと思われる。

3.1 第1,000世代: 記憶長=1の戦略[00]と[01]

第1,000世代では記憶長が1の戦略[00]と[01]がそれぞれ1位と2位の頻度(population内での個体数の割合)を獲得していた。[00]は常にD(積極的)、[01]は前回の相手と逆の行動をする。従って、[00]プレイヤーが[01]プレイヤーと出会うと、

[00] ... D D D D D ...

[01] ... C C C C C ...

のように、「[00]プレイヤーが指導者で[01]プレイヤーが追随者」のように役割が固定化されて、両者の利得の和が最大化される。その意味で、ここでは両者の行動のコーディネーションが実現される。このように、(1)片方のプレイヤーがDを選択肢続けて(2)もう一人のプレイヤーがCを選択し続ける状態、つまり、役割が固定化されて「指導者+追随者」のコーディネーションが実現した状態を、以降、「(固定化した)CD役割分業」と呼ぶことにする¹⁰⁾。

この二戦略によるCD役割分業はエラーに対して頑強である。仮にどちらかのプレイヤーの行動にエラーが発生しても、[00]は自分が間違えたラウンドを除けばDを選択続け、また、[01]はその反対を選択するので、即座にCD役割分業に復帰することができる。

ただし、これら記憶長1の戦略には「同戦略どうしの対戦で行動のコーディネーションができない」という問題がある。例えば、[00] どうしでは両方が共にDを出し続けて譲らないので最悪の結果になる。[01] どうしの場合、前回のお互いの行動の逆を行おうとするので、指導者-追従者関係が一旦形成されればそれは維持されつづける。しかし、エラーが発生すると、次のように、コーディネーションが崩れてしまう。

[01] D D D D C D C D C ...

[01] C C C D C D C D C ...

* (* 印はエラー)

¹⁰⁾Crowley[2]のCorDに相当する。

このように、一旦、エラーで CD 役割分業が崩れると、次のエラーによって CD 役割分業が形成されるまでコーディネーション状態は回復しない。

3.2 第 2,000 世代: 記憶長 2 の戦略 [0100]

第 2,000 世代では、記憶長が 2 の戦略 [0100] の頻度が最も多く、次に [01] の頻度が多かった。まず [0100] のポリシーを見てみよう。

```
-----
1 1 0 0 (f1) 前回の自分 (focal player) の手
1 0 1 0 (o1) 前回の相手 (opponent) の手
[0 1 0 0] 次の手 (=戦略:以下その解釈)
    0    (a) 前回自分が指導者だったら D
    1    (b) 前回相手が指導者なら C
    0    0 (c) 前回行動が衝突したら D
-----
```

[0100] は記憶長が長くなり、前回の相手の行動だけでなく、自分の行動も考慮に入れることができるようになっている。(a) まず、戦略 [0100] の右から 2 番目の「0」は、履歴 (0, 1)₂、つまり前回の自分の行動が D、前回の相手の行動が C の時に次の手は D ということを表している。つまり、前回「自分が追従者で相手が指導者」というコーディネーションができていたら再び指導者を目指す。(b) 次に [0100] の右から 3 番目の「1」は、履歴 (1, 0)₂、つまり前回の自分の行動が C、前回の相手の行動が D の時に次の手は C ということを表している。つまり、前回「自分が指導者で相手が追従者」というコーディネーションができていたら再び指導者を目指す。(c) 最後に、前回の両者の行動が C、あるいは、両者の行動が D の時も指導者を目指して D を出す。[01] は単に前回の相手の行動の逆を選択するが、[0100] が [01] と違う点は、前回両者とも D の場合に D を選択する点である。

[0100] が [01] と対戦では、両プレイヤーの行動が (C, C) なら次は (D, D)、(D, D) なら次は (D, C) になるので、次のように CD 役割分業が実現して、[0100] が指導者、[01] が追従者になる。

```
++ rule (c)
[0100] ... C D D D D D ...
[01] ... C D C C C C ...
```

また、(D, C) 状態でエラーが発生すると必ず (C, C) か (D, D) になるので、その次で CD 役割分業に戻ることができる。つまり、この両者のコーディネーションはエラーに強い。エラーからの回復は次のように行われる。

```
* + + + rule (c)
[0100] D D C D D D D D D D D D ...
[01] C C C D C C C C C D C C C ...
*
```

(+印は上記のルール (c) が適用された場所)

[0100] は [00] のように無条件に D を選択する訳ではなく、CD 役割分業を維持する構造があるので、[00] よりはコーディネーションに成功する。しかし、一旦エラー

が起きて行動の衝突が起きて (D, D) 状態に入ると、次のエラーが起こるまで (D, D) 状態から回復できない。

```
+ rule (c)
[0100] C C C C D D D D ...
[0100] D D D C D D D D ...
* + rule (c)
```

3.3 第 5,000 世代: 記憶長 3 の戦略 [01001100]

第 5,000 世代では、記憶長が 3 の戦略 [01001100] の頻度が最も多く、次に [0100] の頻度が多かった。[01001100] は 10000 世代まで 1 位の頻度を保ち続ける。以下、簡便さのため、この戦略のことを S_1 と呼ぶことにする。この戦略の方針は以下のルール群からなる：(a) 「前回自分が指導者なら D」という記憶長 2 のルール。(b) 「前回相手が指導者なら C」という記憶長 1 のルール。(c) 「前回両者の行動が衝突して、前々回に相手が C だったら D」という記憶長 3 のルール。(d) 「前回両者の行動が衝突して、前々回に相手が D だったら、前回と同じ行動を続ける」という記憶長 3 のルール。

大きく分けると、(a)(b) は一旦 CD 役割分業になった後にコーディネーションを続けるためのルール、(c)(d) は前回コーディネーションに失敗したときのルールである。

記憶長 3 のルールを持つ S_1 は、前々回の記憶まで参照できるので、エラーが発生したときでも、二回前の行動の情報を参照することで CD 役割分業を復活できる。例えば、 S_1 どうしが対戦すると次のようになる。

```
+ rule (d)
S1 C C C C C C ...
S1 D D C D D D ...
* + rule (c)
```

エラーが発生した*印のところで両者の行動が同じになっているが、前々回の行動はお互い異なるので、その情報を参照して CD 役割分業を再開している。

次に、 S_1 が、頻度二位の戦略 [0100] と対戦したとき、 S_1 が指導者の状態からスタートすると次のようになる。

```
+ + rule (c)
S1 ... D D D D D ...
[0100] ... C C D D D ...
* (* はエラー)
```

つまり、 S_1 が指導者になる CD 役割分業はエラーに弱く維持できない。一方、 S_1 が追従者になる CD 役割分業は、次のようにエラーに強い。

```
+ rule (e)
S1 ..C C C C C C ..
[0100] ..D D C D D D ..
*
```

このように、 S_1 は追従者として [0100] と安定した CD 役割分業を実現することができる。

上述のように [0100] どうしが対戦するとエラーによりコーディネーションの失敗が起こるので、 S_1 は、[0100] どうしの時に比べて (指導者にはなれないものの) [0100] に対する better response になる。

以上のように、population で [0100] が増えすぎると S_1 が有利になり、 S_1 が増えすぎると [0100] が指導者になって有利になるという構造があるので、 S_1 と [0100] はタカ・ハトゲームと類似のメカニズムで共存する。ただし、[0100] どちらの対戦ではコーディネーションの失敗が起こりうるのに対し、 S_1 どちらの対戦ではコーディネーションの失敗から回復できるので、population 内での頻度は S_1 が多くなる^{11,12}。

3.4 第4万世代の記憶長4の戦略 [0100 0101 1100 0100]、第8万世代

第4万世代は平均利得が約3.9点と非常に高くなっているが、記憶長が4の戦略 [0100 0101 1100 0100] の頻度が最も多い。以下、簡便さのため、この戦略のことを S_3 と呼ぶことにする。この戦略の方針は以下のルール群からなる：(a)「前回自分が指導者なら D」という記憶長2のルール。(b)「前回相手が指導者なら C」という記憶長1のルール。(c-0)「両者の行動が2回続けて同じなら D」という記憶長4のルール。一種の引き金戦略の性質を備えさせている。(c-1)「前回両者とも C でも、前々回に役割分担が成立していたら、その前々回と同じ行動を行う」というエラーからの回復ルール。(c-2)「前回両者とも D でも、前々回に役割分担が成立していたら、その前々回と逆の行動を行う」という(c-1)と逆方向のエラー回復ルール。

ルール(c-0)は、一種の引き金戦略で、上記(a)(b)(c-1)(c-2)のルールに従わず(C,D)状態が二回連続で成立しないと、その時の相手にはDを出す。ルールを破るとDD状態になるのでルールを破った方も損をする。DD状態になるとD以外出さないの、相手はCを出す以外回復の道はない。

また、ルール(c-1)(c-2)に従っている限り、エラーが起こっても、次のように即座にコーディネーション状態に回復できる。

+ rule (c-2)
 S_3 ..D D D D C C C..
 S_3 ..C C C D D D D..
 * + rule (c-2)

この例のように、ルール(c-2)が適用される時にCとDの役割の交代が起こるため、両プレイヤーは長期的にはほぼ同等の得点を獲得することができる。

自分が追従者の役割に固定されている場合、対戦相手(指導者)が自分と同じ戦略なら、戦略レベルの利得は高くなる。しかし、相手が「似ているけど異なる戦略」の

時には、相手の戦略だけに指導者としての得点を献上してしまうことになる。長期的な役割の交替は、それを防ぐことができる。ただし、この役割交替を悪用される可能性がある。例えば、前々回まで自分が指導者でも、前回、相手が意図的にDを出せば、相手に指導者の立場を奪われる¹³。

3.5 第10万世代 - 指導者ゲームにおける良い戦略 [0#00 #101 0000 0100]

第10万世代での人口頻度上位三戦略は、[0000 1101 0000 0100], [0100 0101 0000 0100], [0000 0101 0000 0100]である。これらの戦略は、#印を0または1として、[0#00 #101 0000 0100]というシンボル列を共通して持っている。

[0#00 #101 0000 0100]は、他の全てのシミュレーション試行における第10万世代の最有力戦略でも共通の部分となっている。このことから、#印以外の部分が特にエラー付き指導者ゲームにおいて重要な部分であることが予想される。簡便さのため、以降、[0#00 #101 0000 0100]のことを $S_{\#}$ と書くことにする。

戦略 $S_{\#}$ の方針は以下のルール群からなる：(a)「両者の行動の衝突が2回続いたら D」という記憶長4のルール。(b-1)「自分が前回に指導者だったら D」という記憶長2のルール。(b-2)「自分が前々回に指導者だったら D」という記憶長4のルール。(c)「相手が二日とも D で、しかも相手が一度以上指導者になっていたら C」という記憶長4のルール。

ルール(a)は、コーディネーションルールを続けて violate する相手には妥協しない、という方策であり、第2万世代、4万世代でも既に現れている。ルール(b1)(b2)は、覚えている範囲(前回と前々回)で自分に指導者の経験があったら積極的にDを選択する、という方策で、指導者の経験(記憶)があると積極的になることを示している。ルール(c)は、ルール(b1)(b2)とは逆に、相手に指導者の経験があるときにどうするかについてのルールであるが、相手がずっと(前回も前々回も)積極的にDを選択している場合にのみ、Cを選択して追従者を目指すということを示している。これにより、例えばエラーによってたまたま指導者になった相手に追従してしまう可能性を排除している。

4 議論

本稿では、エラー付きの繰り返し指導者ゲームにおける戦略の進化をLindgrenモデル[3]を用いて分析した。ここでは、エラー付き指導者ゲーム社会での戦略(ムーアマシン)の進化にはどのような過程がありうるのかと、進化を経た結果としてどのような戦略が指導者ゲームにおいて優勢になっているのかについて議論する。前者については、シミュレーションによって戦略が現れるタイミング等に違いがあるが、本稿で紹介した図1-(a)タイ

¹¹ 以上は直感的な説明だが、数学的には、[0100]と対戦したときの S_1 と [0100] の利得の差 $E(S_1 - [0100] | [0100])$ が、 S_1 と対戦したときの [0100] と S_1 の利得の差 $E([0100] - S_1 | S_1)$ を上回ることにより注意して進化的安定戦略を求めることにより示すことができる。

¹² 以上まとめると次のようになる。第5,000世代では、記憶長3の戦略 $S_1 = [01001100]$ が現れて、[0100]と共にpopulationの多くを占めるようになる。 S_1 は、エラーによるコーディネーションの失敗が起こっても、前々回の相手の行動の情報からCD役割分業を復帰させることができる。ただし、 S_1 は [0100] に対して追従者になるので、完全にpopulationを支配することはできない。

¹³ 紙面の都合で詳細は省略するが、第8万世代には記憶長5の戦略が優勢になる。この記憶長5の戦略は、役割交替のルールを複雑化することで一方的に指導者の立場を奪われるリスクを減らしている。

プの試行の典型例について簡単にまとめる。後者については第10万世代の時点でほぼ共通する構造がみられたのでその構造について議論する。

4.1 戦略進化の経路

上記で紹介したように、エラー付きの繰り返し指導者ゲーム世界において、戦略は、記憶長を徐々に伸ばしたり戦略の内容を修正したりして、過去の戦略の問題点を徐々に修復しながら進化する様子が見られる。

初期の世界では、周りに関係なく常に積極的な戦略と、単に相手の逆の行動を目指すあまのじゃく戦略の共存が見られた。ここでは、積極的な者どうし、あるいは、あまのじゃくどうしが出会ったときに行動のコーディネーションで問題が起こった。しばらく後の世代になると、前々回の相手の行動まで参照することによってコーディネーションの失敗から回復する戦略 (S1) が現れた。しかし、この戦略には相手に合わせすぎて自分が追随者になる頻度が多くなる傾向があった。さらに後の世代になると、前々回の相手と自分の行動まで参照して、連続して役割分業ができなかった相手にはその後は譲らないという一種の引き金戦略のような戦略が現れて、より堅固なコーディネーションを行うようになった。さらに第10万世代までの進化を経た戦略には、次に議論するように $S_{\#}=[0\#00\ 101\ 0000\ 0100]$ という共通のコードが現れることが分かった。

Browning によるエラーなしのケースの研究では、記憶長6のエージェントにより、ステージゲームごとにCとDの役割を入れ替える Alternating reciprocity (AR) が見られた [1]。一方、本研究の結果から分かるように、エラー付き指導者ゲームでは進化の過程でも最後 (10万世代) でも AR は見られず、ほとんどの戦略が固定化した役割分担 (片方のプレイヤーが指導者を続ける) を目指して進化する。

4.2 エラー付き指導者ゲームに適した戦略

最終的に全てのシミュレーションで共通に見られた戦略 $S_{\#}$ は、同じ戦略、同種戦略どうしの間でCD役割分業を行うことが可能になっており、また、エラー等によって立場が変われば同じ戦略でも逆の振る舞いをする。例えば、戦略は全く同じでも、過去の経験によって「積極的な人」と「相手の顔を伺う人」に分岐する。このような振る舞いを可能にする戦略 $S_{\#}=[0\#00\ 101\ 0000\ 0100]$ には以下の3つのポリシーがある。

まず一つめは「上手く行動をコーディネートできない相手に、自分からは譲らない」というポリシーである。両者が譲らなくなると、どちらかのプレイヤーが (エラーによって、あるいは、戦略的に) 譲るまで、最悪の状態が続くことになる。コーディネーションが成立しなくなった時をきっかけに攻撃的な戦略を選ぶ、という点で、繰り返し囚人ジレンマの引き金戦略と類似点がある。

二つめは、「いったん指導者の経験をすると積極的に振る舞う」というポリシーである。ノイズなしのケースに

関する研究 [1] では、指導者になると次のステージゲームでは消極的になるので対照的な結果である。

三つ目は、「自分に指導者になった記憶がない場合、相手が常に積極的で指導者の経験もあるなら相手に追従する」というポリシーである。つまり、自分に指導者の経験がない時には、まず相手の様子をうかがい、相手が確実に指導者であることを確認したら (つまり、相手が指導者の場合も、偶然指導者になったのではなく、普段から積極的であることを確認したら) 相手に追従する。

以上のようなポリシーで行動する実際の人間がどの程度いるのかは、厳密には被験者実験等で検証する必要があるが、本研究が示唆するのは、ノイズ付き指導者ゲームのような相互作用がある場合、進化の結果として、以上のようなポリシー・戦略が生き残る可能性があるということである¹⁴。

参考文献

- [1] Lindsay Browning and Andrew M. Colman. Evolution of coordinated alternating reciprocity in repeated dyadic games. *Journal of Theoretical Biology*, Vol. 229, No. 4, pp. 549–557, August 2004.
- [2] Philip H. Crowley. Dangerous games and the emergence of social structure: evolving memory-based strategies for the generalized hawk-dove game. *Behav. Ecol.*, Vol. 12, No. 6, pp. 753–760, November 2001.
- [3] K. Lindgren. Evolutionary phenomena in simple dynamics. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pp. 295–312. Addison Wesley Publishing Company, 1991.
- [4] M. Nowak and K. Sigmund. Chaos and the evolution of cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 90, No. 11, pp. 5091–5094, June 1993.
- [5] Anatol Rapoport. Exploiter, leader, hero, and martyr: The four archetypes of the 2x2 game. *Behavioral Science*, Vol. 12, No. 2, pp. 81–84, 1967.
- [6] Jun Tanimoto and Hiroki Sagara. A study on emergence of alternating reciprocity in a 2x2 game with 2-length memory strategy. *Biosystems*, Vol. 90, No. 3, pp. 728–737, 2007.
- [7] J. Wu and R. Axelrod. How to cope with noise in the iterated prisoner's dilemma. *Journal of Conflict resolution*, Vol. 39, pp. 183–189, 1995.

¹⁴今後の課題の一つとして、世代数とシミュレーションの試行回数を増やして、本研究の結果の頑健性をチェックする必要がある。