

有限オートマトンを用いた私的観測繰り返しゲーム における進化的安定戦略分析

小池 淳平¹ 澤 亮治^{2,a)}

受付日 2021年3月3日, 採録日 2021年12月3日

概要: 私的観測繰り返し囚人のジレンマゲームにおける進化的安定戦略の分析を行った。私的観測とは、相手の行動を見ることができない代わりに相手の行動に依存する不確実なシグナルを受け取る状況を指す。完全観測の繰り返しゲームにおける有限オートマトン戦略の進化的安定性の分析を行った Binmore and Samuelson (1992, Journal of Economic Theory) の定義を私的観測に拡張し、数値解析により進化的安定な (2 状態以下の) 有限オートマトン戦略を判定した。その結果、4 種類の有限オートマトンが進化的に安定となることが判明した。この 4 種類の有限オートマトンが実際に出現しやすいかを進化シミュレーションにより追加検証した。

キーワード: 有限オートマトン, 私的観測, 繰り返しゲーム, 進化的安定性

An Analysis on Evolutionary Stability of Strategies Using Finite Automata in Repeated Games with Private Monitoring

JUMPEI KOIKE¹ RYOJI SAWA^{2,a)}

Received: March 3, 2021, Accepted: December 3, 2021

Abstract: Evolutionarily stable strategies in repeated prisoner's dilemma games with private monitoring are examined. Private monitoring refers to settings where players can not observe other players' actions perfectly, but instead receive noisy signals about their actions. We extend the definition of evolutionary stability in Binmore and Samuelson (1992, Journal of Economic Theory) to repeated games with private monitoring, and examine evolutionary stability of (up to two-state) finite automata via a numerical analysis. We find that four types of automata are evolutionarily stable. We further conduct evolutionary simulations to examine whether these automata emerge in such environments.

Keywords: finite automata, private monitoring, repeated game, evolutionarily stable strategy

1. はじめに

繰り返しゲームの理論は、プレイヤー間に長期的なやり取りが存在することで構築できる関係性を明らかにする。短期的には不可能な協力関係が長期的な関係では構築可能となることなどが明らかになっている。同一のゲームを無限に繰り返すゲームを無限繰り返しゲームと呼び、一般的に

このようなゲームは無数の均衡を持つ (文献 [1] の 3.2 節や文献 [2] の 2.4-2.5 節など参照)。均衡が複数あるゲームにおいて出現しやすい均衡を選択する手法を均衡の精緻化と呼ぶ。進化ゲーム手法は均衡の局所のおよび大域的な安定性に着目し、均衡の精緻化を行う。本研究では有限オートマトンで表現可能な戦略に注目し、進化的手法による私的観測繰り返し囚人のジレンマゲームの均衡の精緻化を行う。

繰り返しゲームには多数の戦略が存在するが、2 状態以下の有限オートマトンで表現可能な戦略のみを分析対象とする*1。プレイヤーの戦略に制限を設けるが、繰り返しゲームの経済実験では 2 状態の有限オートマトンで表現可能な

¹ 株式会社大和総研
Daiwa Institute of Research Ltd., Koto, Tokyo 135-8460, Japan

² 筑波大学システム情報系社会学工学科
Graduate School of Systems and Information Engineering,
University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

a) rsawa@sk.tsukuba.ac.jp

*1 戦略のオートマトン表現は、たとえば文献 [3] に簡単な解説がある。

戦略が実際によく観察されており、一定の現実性を持つ戦略集合である (文献 [4] など)。また、繰り返しゲームの中でも「協力したのに相手に誤解された」「裏切ったことが相手にバレなかった」といったお互いの行動が完全に観測できない不完全観測の状況を扱う。特に、不完全観測において「相手の行動を不完全に表すシグナル」が個別に観察できる私的観測の状況を分析対象とする。

上記の私的観測の繰り返しゲームにおける 2 状態以下の有限オートマトン戦略の安定性を 2 つの精緻化手法により検証する。第 1 に、進化的安定戦略 (Evolutionarily Stable Strategy, 以下 ESS) の概念を用いて局所的に安定な戦略を同定する。ESS の戦略を全プレイヤーが採用している状態では他の戦略が侵入できないという種類の局所的安定性を ESS は持つ (文献 [5])。第 2 に、文献 [6] の進化生態学的シミュレーションを用い、ESS の戦略が採用される均衡が大域的に出現しやすいかの検証を行う。各戦略を採用しているプレイヤー数が均等である初期状態から、プレイヤーが徐々に戦略を変更していく遷移を進化動学のシミュレーションにより分析する。均衡からの乖離が大きい状態から ESS の戦略を用いる均衡への収束を検証するという点で、大域的安定性を確認する手法である*2。ESS という局所的安定性を持つ戦略が、実は大域的な安定性を持たず進化シミュレーションなどで出現しにくいことがある (文献 [7])。そのため、数値解析による ESS 判定と進化シミュレーションの 2 つの手法を併用し安定性を検証する。

本研究の貢献は、私的観測囚人のジレンマ繰り返しゲームにおいて (i) 様々なエラー確率 (相手の行動を見誤る確率) における ESS 戦略の数値解析による同定、および (ii) 進化シミュレーションによる ESS 戦略の大域的安定性の検証の 2 点である。(i) の数値解析によりエラー確率と ESS 戦略の関係を示す。低いエラー率では相手の裏切り行動 (D) を観察した場合に行動を切り替える Pavlov 戦略が ESS となり、中程度のエラー率では 1 度でも裏切り行動を観察したら永久に裏切り行動を続けるトリガー戦略が ESS となる。つねに裏切り行動をとる All-D 戦略はすべてのエラー率で ESS となる。また、(ii) のシミュレーションにより、ESS の点では最も安定的である All-D 戦略が必ずしも大域的安定性を持たないことを示す。エラー率が低いケースでは、Pavlov 戦略またはトリガー戦略への進化シミュレーションの収束が多くの場合に確認された。

有限オートマトンを用いた囚人のジレンマの繰り返しゲーム (以降、マシニングゲーム) について ESS など局所的安定性の観点から行われた関連研究を述べる*3。文献 [8], [9] はマシ

ニングゲームのナッシュ均衡の分析を行った。文献 [10], [11], [12] は有限オートマトンの複雑さを考慮した ESS を定義し、マシニングゲームの分析を行った。文献 [10] では、第 1 に利得、第 2 に複雑さの少なさを優先する辞書式選好を導入した ESS が提案されている。文献 [11], [12] も異なる複雑さの導入の仕方により ESS の定義を行っている。また、文献 [13], [14] は主に文献 [12] の定義を用いて ESS となる戦略を分析した。上記研究は完全観測の状況を対象としており、本研究は文献 [10] の ESS 定義を援用して不完全観測の状況の分析を行った。

不完全観測の状況については、文献 [17] がエラーを含む有限オートマトン戦略の安定性を分析している*4。エラー率を 0 に近づけた極限の利得では、All-D 戦略、トリガー戦略、Pavlov 戦略が広い利得の範囲で駆逐されない戦略であることを示した*5。ただし、文献 [17] では有限オートマトンの状態を直接遷移させるエラーが起こりうる。本研究ではエラーによる状態遷移は相手の行動を見間違える観測エラーを介した遷移のみを仮定している。文献 [18] は、2 状態以下の有限オートマトン戦略を対象にし、本研究と同様に相手の行動を誤って観測するタイプのエラーを仮定している。エラー確率を 0 に近づけた極限の利得を用い、一定の利得条件の下で Pavlov 戦略のみが進化的に安定となることを示した。本研究では、利得設定は限定し、エラー率の変動による進化的安定性への影響を検証している。

進化動学を利用した大域的な安定性の検証については、囚人のジレンマ繰り返しゲームのコンピュータトーナメント (文献 [6], [19]) の参加プログラムによる進化シミュレーションを Axelrod [6] が実施している。最初に協利行動 (C)、その後は前期の相手の行動を模倣するしつぺ返し戦略が、進化シミュレーションでの生き残り戦略であった。しかし、つねに協利行動を行う戦略 (All-C 戦略) としつぺ返し戦略が対戦した場合には双方同じ利得となり、かつ All-C 戦略は 1 状態有限オートマトンで表現可能なより単純な戦略である。そのため、しつぺ返し戦略は ESS とはならない。このように ESS と進化シミュレーションの結果が異なる場合もあり、均衡から離れた状態からの進化動学による検証など大域的な安定性の議論も重要となる。

不完全観測のマシニングゲームの進化動学の分析については、文献 [20], [21], [22] などがある。文献 [20] は 2 状態以下の有限オートマトンを戦略としたレプリケータダイナミクスからの帰結から、出現しやすい戦略を分析した。文献 [20] はゲームの利得構造と出現する戦略の関係に焦点を当てており、本研究はシグナル観測のエラー率と割引因子のパ

*2 ここでの大域的安定性は、各戦略のプレイヤー数が均等である初期状態からの収束性という狭い範囲での安定性を意味している。

*3 プレイヤーが同時に行動を決定する戦略形ゲームを繰り返す状況に対する有限オートマトン戦略の研究をあげる。プレイヤーが順番に行動を決定する展開形ゲームを繰り返す状況に関する研究は、文献 [7], [15], [16] などがある。

*4 以降で紹介する研究の一部は記憶容量 1 の戦略 (one memory strategy) を対象としている。記憶容量 1 の戦略と 2 状態以下有限オートマトンの対応については文献 [17] の 2 章が詳しい。

*5 駆逐されない戦略の定義は進化的安定性より緩い。厳密に正確な表現ではないが、すべての他の戦略 s に対して (i) s に侵入されないか、(ii) s に侵入できるかのいずれかが成り立つ。

ラメータ変化が均衡の安定性に及ぼす影響の考察が主である。文献 [20] については、4.5 節でより詳しく議論をする。文献 [21] は本研究と補完関係にあり、様々なエラー率における損失回避性の有無による進化的安定性への影響を分析している。文献 [22] は 2 人のプレイヤーがお互いの見間違えをほぼ共有する状況を想定しており、本研究の独立した見間違いと対極にある状況といえる。このような状況では Pavlov 戦略が生き残り戦略とならないことが報告されている。

上記以外にも、エラーを考慮したマシニングゲームの進化動学分析には文献 [17], [18], [23], [24], [25], [26] などの関連研究がある。文献 [23] は 0.1% 以上の確率で行動を誤る有限オートマトン戦略を用いて進化シミュレーションを行い、Pavlov 戦略が生き残り戦略となることを示した。ただし、文献 [17] と同様に有限オートマトンの状態を直接遷移させるエラーを仮定している。文献 [17] は文献 [19] の利得設定においてエラー率 0 の極限で計算した利得を用いて、進化動学を分析した。この利得設定では 3 戦略以上の多くの組合せで動学が収束しないサイクルが発生する。文献 [24] は評判システムを仮定した状況で、一定確率で行動を誤る有限オートマトン戦略を用いた進化シミュレーションを行った。評判を考慮した形で改善したしっぺ返し戦略が生き残り戦略となりうることが示された。文献 [18] は 3 種類のエラー確率 (0%, 1%, 4%) において遺伝的アルゴリズムを利用した進化シミュレーションを行った。平均利得と閾値の比較から行動を選ぶ戦略を遺伝子として表現し、低エラー率であるほど協力が高まったことが報告されている。文献 [25] はエラーを含む 4 つの戦略 (All-C, All-D, Pavlov, しっぺ返し) について、プレイヤーの行動が均衡に収束した後も非常に低確率の突然変異により均衡間の遷移が起こる進化動学を分析した。均衡間の遷移を考慮した超長期的な動学では、All-D または Pavlov が出現しやすい。本研究は均衡間の遷移までは考慮しない中期的な収束に着目している。文献 [26] は展開形ゲームに変形した囚人のジレンマを繰り返すゲームを分析し、All-D, Pavlov, トリガー、しっぺ返しとは異なる戦略がエラーありの環境で生き残ることを示した。

私的観測繰り返しゲームの理論研究に信念ベースアプローチと呼ばれる手法がある^{*6}。相当に砕けた表現をすると、この手法は均衡どおりの行動をとるインセンティブがある均衡の同定手法であり、進化的安定性の考えと近い。文献 [27] は、強化学習にも用いられる部分観測マルコフ決定過程を利用し、有限オートマトンが私的観測繰り返しゲームの逐次均衡となる判定法を示した。文献 [28] は文献 [27] の考えを用いて 3 状態以下の有限オートマトンで構

成される均衡を数値分析した。コンピュータを利用した数値分析という手法は類似しているが、文献 [28] では高い利得を実現する均衡に着目しており、本研究では均衡の安定性に主眼を置いている。

2. モデル

2.1 私的観測繰り返し囚人のジレンマゲーム

ゲームの定式化を行う。プレイヤー集合を $N = \{1, 2\}$ 、プレイヤー $i \in N$ の行動集合を $A_i = \{C, D\}$ 、プレイヤー i が受け取るシグナルの集合を $\Omega_i = \{g, b\}$ とする。 Ω_i は、プレイヤー i が相手の行動を観測する代わりに受け取るシグナル ω_i の集合である。2 人のプレイヤーの行動のペア $\mathbf{a} = (a_1, a_2)$ を行動プロファイルと呼び、2 人が受け取るシグナルのペア $\boldsymbol{\omega} = (\omega_1, \omega_2)$ をシグナルプロファイルと呼ぶ。シグナルプロファイルは行動プロファイルに依存して確率的に決定され、 $o(\boldsymbol{\omega}|\mathbf{a})$ をその (同時) 確率分布とする。観測にエラーがない場合には、プレイヤー $i \neq j \in N$ について、 $a_j = C$ の場合には $\omega_i = g$ 、 $a_j = D$ の場合には $\omega_i = b$ となるとする。また、プレイヤー i の利得は自身の行動 a_i と観測したシグナル ω_i にのみ依存する。関数 $\pi_i : A_i \times \Omega_i \rightarrow \mathbb{R}$ を利得関数とし、各 $i \in N$ について以下の利得を仮定する。

$$\begin{aligned} \pi_i(C, g) &= 3, & \pi_i(C, b) &= 0, \\ \pi_i(D, g) &= 4, & \pi_i(D, b) &= 1. \end{aligned} \tag{1}$$

完全観測の下で進化的に安定な戦略との比較を容易にするため、文献 [10] と同様の利得設定とした^{*7}。ゲームは囚人のジレンマの構造を持つ。どのシグナルに対しても、 D がつねに利得を最大化する最適な行動となる。しかし、(シグナルのエラーが小さければ) お互いが最適な行動 D を選んだ際に得られる利得はおおよそ 1 となり、お互いが C を選んだ際に得られる利得よりも小さい。また、 $2\pi_i(C, g) > \pi_i(D, g) + \pi_i(C, b)$ であり、 (C, C) を続ける方が $(D, C), (C, D)$ を繰り返すよりも双方にとって良い。

行動プロファイル \mathbf{a} のときのプレイヤー i の期待利得は

$$g_i(\mathbf{a}) = \sum_{\boldsymbol{\omega} \in \Omega_1 \times \Omega_2} \pi_i(a_i, \omega_i) o(\boldsymbol{\omega}|\mathbf{a})$$

と表せる。

シグナル $\boldsymbol{\omega}$ の同時分布については、分析を単純にするため、誤ったシグナルが発生するエラー率を p とした互いに独立したシグナルを各プレイヤーが受け取るとする。つまり、 $a_j = C$ の場合には $\omega_i = b$ 、 $a_j = D$ の場合には $\omega_i = g$ となる事象が確率 p で発生する。例として、行動プロファイル $\mathbf{a} = (C, C)$ のシグナル確率を以下に示す。

$$o((g, g)|(C, C)) = (1 - p)^2$$

$$o((b, g)|(C, C)) = p(1 - p)$$

^{*6} 他にコミュニケーションアプローチ (文献 [29], [30] など) や、信念不問アプローチ (文献 [31], [32], [33] など) がある。概要は文献 [2] を参照。

^{*7} 4 章の分析の都合上、数値が非負となるように文献 [10] のゲームの利得へ 1 を足した値とした。

$$o((g, b)|(C, C)) = (1 - p)p$$

$$o((b, b)|(C, C)) = p^2$$

本研究では、両プレイヤーが相関のない独立したシグナルを受け取る場合を検討する。お互いが共通の手法を用いずに相手の行動を観測しようとする場合などが該当する。相関を仮定した場合には、自分が受け取ったシグナルから相手のシグナルの推測も可能となるが、このような推測に関する戦略性を分析に含めず、誤りを含むシグナルに対する反応とその帰結を分析のメイン対象とする。また2人の受け取るシグナルの精度 $1 - p$ は差がないとした。受け取るシグナルの精度に差がある状況の分析は今後の課題である。

プレイヤーは上記のゲームを無限回繰り返す。プレイヤー i の $t \in \{1, 2, \dots\}$ 期までの過去の行動と受けとったシグナルの記録を i の私的履歴と呼び、 $\mathbf{h}_i^t = (a_i^1, \omega_i^1, \dots, a_i^t, \omega_i^t) \in H_i^t := (A_i \times \Omega_i)^t$ で表す。ここで、 a_i^s はプレイヤー i の s 期の行動を、 ω_i^s は i が s 期に受け取ったシグナルを示す。無限回繰り返しゲームにおけるプレイヤーの純粋戦略 s_i は、 $H_i = \bigcup_{t \geq 1} H_i^t$ と置くと、 $s_i := H_i \rightarrow A_i$ と表せる。ここで、 $t + 1$ 期の行動は $s_i(\mathbf{h}_i^t)$ で与えられる。割引因子 $\delta \in (0, 1)$ とし、無限繰り返しゲームの期待利得を以下のように表す (\mathbf{h}_i^0 は空集合とする)。

$$f_i(s_1, s_2) = (1 - \delta)E \left[\sum_{t=0}^{\infty} \delta^{t-1} g_i(s_1(\mathbf{h}_1^t), s_2(\mathbf{h}_2^t)) \right]$$

2.2 純粋戦略の有限オートマトン表現

本研究では、2.1 節で述べた私的観測繰り返しゲームをプレイする状況を考えるが、プレイヤーの戦略を2状態以下の有限オートマトンで表現できる戦略に限定する。以下では、有限オートマトンで表現される戦略をマシンと呼び、戦略を有限オートマトンに限定したゲームをマシンゲームと呼ぶ。オートマトンは各状態での行動、および相手から受け取るシグナルに対する状態遷移の規則を定め、毎期にとる行動を決める。プレイヤー i の用いるマシンを m_i とし、 $m_i = \langle Q_i, q_i^0, \lambda_i, \mu_i \rangle$ の4つの要素で表す。 Q_i はマシン m_i における有限の状態集合、 q_i^0 はマシン m_i の初期状態 ($q_i^0 \in Q_i$) である。 $\lambda_i : Q_i \rightarrow A_i$ は行動決定関数であり、状態 $q_i \in Q_i$ から行動 $a_i \in A_i$ を返す。 $\mu_i : Q_i \times \Omega_i \rightarrow Q_i$ は状態遷移関数であり、自分の状態 $q_i \in Q_i$ と受け取ったシグナル $\omega_i \in \Omega_i$ から次の期の自身の状態 $q_i \in Q_i$ を返す。マシンゲームにおける利得は以下のように表せる。

$$P_i(m_1, m_2) = (1 - \delta)E \left[\sum_{t=0}^{\infty} \delta^{t-1} g_i(\lambda_1(q_1^t), \lambda_2(q_2^t)) \right] \quad (2)$$

ここで、 q_i^t は t 期でのプレイヤー i のマシンの状態を示す。マシン m_i の複雑さを状態数の大きさ $|Q_i|$ と定義する。 M を状態数が2以下のマシンの集合とする。つまり

$M = \{m = \langle Q_i, q_i^0, \lambda_i, \mu_i \rangle : |Q_i| \leq 2\}$ である。以下の分析では、 M が各プレイヤーの戦略集合となる。また、以降では便宜的にマシン $m \in M$ の複雑さを $|m|$ と表す。

3. 進化的安定性の数値解析判定

3.1 進化的安定戦略

不完全観測のマシンゲームへ拡張した進化的安定戦略の定義を述べる。文献 [10] と同様に、利得を第1、複雑さを第2の優先順位として進化的安定戦略 (ESS) を定義する。

定義 3.1 $m^* \in M$ が進化的安定戦略であるとは、 $\forall m \neq m^* \in M$ に対して以下の (I), (II), (III) の条件のいずれかが成り立つことをいう。

- (I) $P_1(m^*, m^*) > P_1(m, m^*)$,
- (II) $P_1(m^*, m^*) = P_1(m, m^*)$ and $P_1(m^*, m) > P_1(m, m)$,
- (III) $P_1(m^*, m^*) = P_1(m, m^*)$, $P_1(m^*, m) = P_1(m, m)$, and $|m^*| < |m|$

利得 P_1 が完全観測繰り返しゲームのマシンゲームの利得であれば、この定義は文献 [10] と一致する。条件 (I) はマシン m^* どちらの状態から別のマシン m へ逸脱すると利得が減少することを意味する。条件 (II) はマシン m は m^* との対戦から得る利得がマシン m^* どちらが対戦した利得と等しいが、マシン m どちらの対戦から得る利得は m^* が m の対戦から得る利得よりも低いことを意味する。条件 (I) または (II) が成り立つ場合には、すべてのプレイヤーが m^* を採用している状況で m へ逸脱した (複数かつ少数の) プレイヤーの利得は必ず他のプレイヤーより低くなる。条件 (III) は条件 (II) までの利得がそれぞれ同じとき、 m^* の状態数は m の状態数より小さくしなければならないことを意味する。マシンの状態数が多いほど複雑な戦略が表現可能となる。戦略の実行には複雑さに比例した非常に微小なコストがかかり、利得が同じであれば複雑さの少ない側が有利との想定をした条件となる。また、マシンゲームには定義 3.1 を満たす戦略は存在しない状況 (文献 [10], 割引因子 $\delta = 1$ の場合など) があるため、文献 [10] は (III) の条件を以下の (IV) へ緩めた安定性も定義した。

定義 3.2 $m^* \in M$ が修正進化的安定戦略 (Modified Evolutionarily Stable Strategy, 以下 MESS) であるとは、 $\forall m \neq m^* \in M$ に対して (I), (II), (IV) のいずれかの条件が成り立つことをいう。

- (IV) $P_1(m^*, m^*) = P_1(m, m^*)$, $P_1(m^*, m) = P_1(m, m)$ and $|m^*| \leq |m|$

定義より自明であるが、ESS の条件を満たすマシンは MESS の条件も満たす。本研究では定義 3.1 の ESS を用いて、進化的安定性の判定を行った。完全観測状況との比較には、文献 [10] で報告されている MESS の戦略を用いる。

3.2 ESS 判定手法

数値解析による ESS 判定手法を述べる。繰り返しゲーム

表 1 割引因子 $\delta = 0.99$ における ESS 判定結果 (%はエラー率)

Table 1 ESS identification result for the discount factor $\delta = 0.99$ (The numbers with % denote the error rates).

マシン番号	○	△	×
3 (Pavlov-C)	1~4%	5~6%	7~30%
4 (トリガー)	3~16%	1~2%, 17~19%	20~30%
16 (Pavlov-D)	1~6%		7~30%
26 (All-D)	1~30%		

(○: 9セット以上で ESS 判定, △: 1~8セットで ESS 判定, ×: 0セットで ESS 判定)

表 2 割引因子 $\delta = 0.95$ における ESS 判定結果 (%はエラー率)

Table 2 ESS identification result for the discount factor $\delta = 0.95$ (The numbers with % denote the error rates).

マシン番号	○	△	×
3 (Pavlov-C)	1~6%		7~30%
4 (トリガー)	5~17%	2~4%, 18~19%	1%, 20~30%
16 (Pavlov-D)	1~6%		7~30%
26 (All-D)	1~30%		

(○: 9セット以上で ESS 判定, △: 1~8セットで ESS 判定, ×: 0セットで ESS 判定)

の最終期は $T = 1,000$ とした。つまり、式 (2) の無限回 ∞ を $1,000$ で置き換えた利得を判定に使用する。観測ノイズにより利得が確率的となるため、 $T = 1,000$ のマシンゲームを $1,000$ 回行い、 $1,000$ 回の利得の平均値を期待利得 P_i として用いた*8。割引因子 $\delta = 0.99, 0.95$ およびエラー率 $1\sim 30\%$ (1%ごと) の各パラメータ設定で定義 3.1 による判定を行った。利得の不確実性による判定誤差の可能性を考慮し、 $T = 1,000$ のマシンゲーム $1,000$ 回分の平均利得を計算する数値解析を計 10 セット行った。10 セットの数値解析において、ESS 判定を 9 セット以上で得た場合に達成 (表では ○)、1~8 セットで不確実に達成 (表では △)、0 セットで未達成 (表では ×) とした。付録 A の図 A-1 に 2 状態以下の有限オートマトンをすべて示す。以降の番号付きのマシン名は図内の有限オートマトンの番号と対応している。いくつかのマシンにはその特徴を示す名前を与え、図 A-1 の各マシンの欄右下に名前を記した (たとえば、マシン 13 はつねに行動 C を選ぶ戦略であり、All-C 戦略と名付けた)。

3.3 数値解析結果

数値解析の結果、4 種類の有限オートマトンが ESS となることが判明した。ESS となるオートマトンは割引因子およびエラー率に依存する。表 1 に割引因子 $\delta = 0.99$ の際の各エラー率における判定結果を示す。同様に割引因子 $\delta = 0.95$ の結果を表 2 に示す。ESS と判定した有限オートマトンは図 1 に示す。表内のマシン番号は図 1 に示した各オートマトンの番号と対応している。

図 1 の各オートマトンの円は状態を示し、円の状態にあるときは円内に示された行動を選ぶ。2 状態オートマトンでは左側の円で示される状態が初期状態となる。矢印は各シグナル g, b を受け取った際の状態遷移を示す。以下で各オートマトンについて説明する。マシン 3 は初期状態 C で、受け取ったシグナルが g ならば状態 C 、 b ならば状態

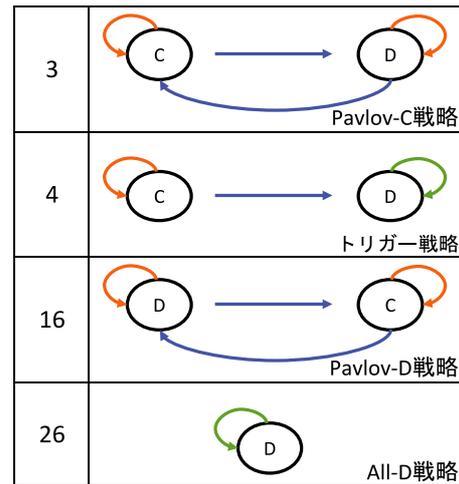


図 1 ESS となる有限オートマトン戦略 (各矢印は橙がシグナル g 観察時の遷移、青はシグナル b 観察時の遷移、緑はシグナル g または b 観察時の遷移を表す)

Fig. 1 Finite automata identified as ESS (Orange arrows show transitions when signal g is observed, blue arrows show transitions when signal b is observed, and green arrows show transitions when either signal g or b is observed).

D に遷移する。状態 D のとき、受け取ったシグナルが g ならば状態 D 、 b ならば状態 C に遷移する。マシン 4 は初期状態 C で、受け取ったシグナルが g ならば状態 C 、 b ならば状態 D に遷移する。また状態 D に遷移後はどのシグナルを受け取っても状態 D にとどまり続ける。シグナル b の受け取りがトリガーとなり、永遠に状態 D となるためトリガー戦略と呼ばれる。マシン 16 はマシン 3 とまったく同じ状態遷移を持つが、初期状態が D であることが異なる。マシン 3 と 16 は、良い結果 (シグナル g) であれば現在の行動を維持し、悪い結果 (シグナル b) であれば行動を変更する条件付きの反応を示す戦略と解釈でき、Pavlov 戦略とも呼ばれる (文献 [34])。ここでは、マシン 3 を Pavlov-C 戦略、16 を Pavlov-D 戦略と呼ぶ。マシン 26 はつねに行動 D をとり続ける戦略であり、All-D 戦略と呼ぶ。

3.4 考察

割引因子 $\delta = 0.99$ において、同じマシンどうしが対戦したときのエラー率 (横軸) と利得 (縦軸) の関係を図 2 の

8 式 (2) より、割引因子 $\delta = 0.99$ における $1,000$ 期時点の利得は $(1 - 0.99) \times 0.99^{1000} \approx 4.3 \times 10^{-7}$ の重みを持つ。ESS 判定を得たマシンは条件 (I) の別のマシン m への逸脱の際の逸失利得 $(P_1(m^, m^*) - P_1(m, m^*))$ の最小値は小数点 3 桁以上の数値 (おおよそ 0.02 以上) となっていた。1,001 期以降の利得が与える逸失利得への影響は小さく、1,000 期分の数値解析で判定可能と判断した。

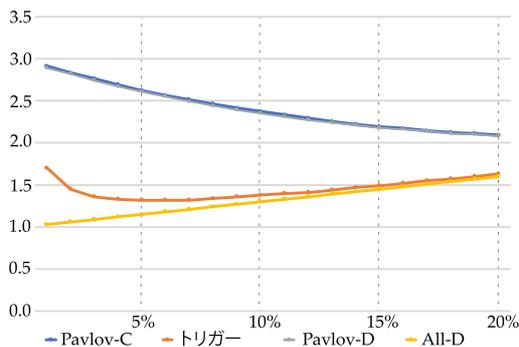


図 2 割引因子 $\delta = 0.99$ の利得変化 (横軸: エラー率, 縦軸: 利得)
 Fig. 2 Relationship between payoffs and error rates for discount factor $\delta = 0.99$ (x-axis: error rate, y-axis: payoff).

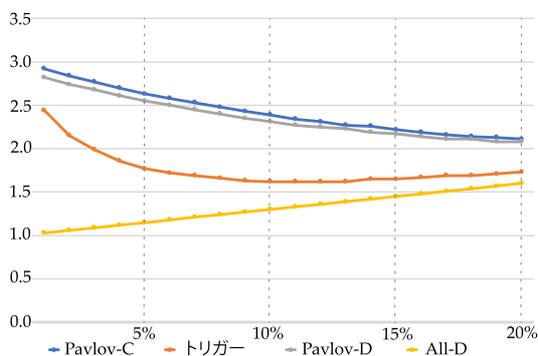


図 3 割引因子 $\delta = 0.95$ の利得変化 (横軸: エラー率, 縦軸: 利得)
 Fig. 3 Relationship between payoffs and error rates for discount factor $\delta = 0.95$ (x-axis: error rate, y-axis: payoff).

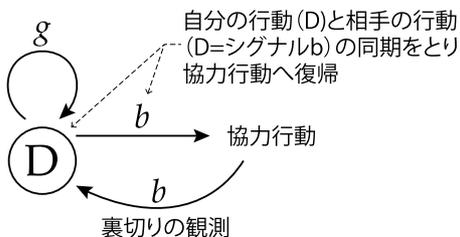


図 4 協力復帰の鍵となるパーツ

Fig. 4 A key component of recovery to cooperation.

グラフに示す。 $\delta = 0.95$ における同様のグラフを図 3 に示す。図 2, 図 3 より, 低いエラー率において, Pavlov-C 戦略と Pavlov-D 戦略は高い利得と ESS 達成を両立していることが分かる。これらのマシンは, 悪いシグナル b を受け取っても協力行動に復帰できる性質を持つ。この仕組みの鍵となるパーツを図 4 に示す。Pavlov-C および Pavlov-D 戦略はこのパーツを持つ。たとえば, Pavlov-C 戦略どうしの対戦で, お互いに協力をしている状態 (状態 C) から b のシグナルをプレイヤー 1 が受け取り, 状態 D に遷移したとする。その後エラーが起こらなかったとすると, b を観測したプレイヤー 1 は次の期に D 行動をとり, プレイヤ 2 は C 行動をとる。プレイヤー 2 は, 1 の D 行動から b を受

け取り, 状態 D へ遷移する。プレイヤー 1 は, g を受け取るため, D 状態のままである。その次の期にお互いに行動 D をとり, 双方が b を受け取り, 同時に状態 C へ遷移する。つまり, エラー後の行動プロファイルは (D, C) , (D, D) , (C, C) と遷移し協力行動に復帰することができる。また, お互いに協力をしている状態 (状態 C) から b のシグナルを両方のプレイヤーが受け取り, 両方が状態 D に遷移した場合には, エラー後の行動プロファイルは (D, D) , (C, C) となり, やはり協力行動に復帰可能である。

Pavlov-C および Pavlov-D 戦略は低いエラー率では ESS となるが, エラー率が大きくなると ESS とならない。協力行動に復帰する性質を持っていても, エラー率が高いとノイズにより協力行動に復帰後に (C, b) の利得が実現し, D 状態に移りやすくなる。そのため, 行動プロファイル $a = (C, C)$ を長く継続できず, 同じ戦略どうしの対戦の利得が低くなる (図 2, 図 3)。一方で, 協力行動に復帰する性質は All-D 戦略など D 行動を続ける戦略に搾取される傾向がある。たとえば, Pavlov-C 戦略であれば, All-D 戦略の対戦では (C, D) , (D, D) , (C, D) , (D, D) の遷移が続き, (割引因子によらず) 自身がおおよそ 0.5 の平均利得を相手がおおよそ 2.5 の平均利得を得る。このときの相手が得る利得はエラー率によらずほとんど変わらない (自分の C 行動時に b が発生する確率と D 行動時に g が発生する確率が等しいためである)。そのため, 自身の戦略どうしの対戦から得られる利得がおおよそ 2.5 を下回るエラー率で ESS の条件を満たさなくなる。これが起こるのが, 6~7%のエラー率である (図 2, 図 3)。Pavlov-C 戦略は, エラー率が高くなるとマシン 12 との対戦で定義 3.1 の条件 (I) を満たさない。マシン 12 は初回のみ C 行動を選び, その後はシグナルにかかわらず D 行動を選ぶという All-D 戦略に類似した戦略である。Pavlov-D 戦略は, エラー率が高くなると All-D 戦略との対戦で定義 3.1 の条件 (I) を満たさない。

エラー率が比較的高い範囲では, シグナル b を受け取り後は永遠に行動 D をとるという裏切りへの厳罰を行うトリガー戦略が ESS を達成する。トリガー戦略どうしの対戦では初期は (C, C) の協調行動により比較的高い利得が得られる。エラー発生後は (D, D) が続く状態に移行するが, エラー率が高くなると行動プロファイル (D, D) を取っていてもエラーによって利得 $\pi_i(D, g) = 4$ を得られる確率が高い。図 2, 図 3 より, 中程度のエラー率よりエラー率上昇に対して利得が増加傾向にあることが分かる。このような利得傾向から比較的高いエラー率で ESS を達成しやすくなっていると考えられる。

トリガー戦略は割引因子 0.99 における 1~2%, および割引因子 0.95 における 1~4% など, 低いエラー率で ESS となりにくい傾向がある。非常に低いエラー率でトリガー戦略が ESS とならないのは, All-C 戦略との対戦で定義 3.1

の条件 (I) を満たさないことが原因である。非常に低いエラー率では b を観察後に D 状態に移ることによる逸失利得が大きくなる。 b を観察した場合に D の罰状態への移行を仮に見逃したとすると、もう 1 度エラーが起こるまでは (C, C) を継続することができる。低いエラー率ではこのような見逃しの方が利得が高まる場合があり、そのような見逃しを行う All-C 戦略がトリガー戦略との対戦で得る利得がトリガー戦略どうしの利得を上回る。

All-D 戦略はエラー率によらずつねに ESS となった。All-D 戦略が進化的安定という結果は、(進化的安定性の定義は異なるが) 文献 [11] や 1 より小さい割引因子を考慮した文献 [14] の結果と整合的である。特に高いエラー率では、 (C, C) の協調行動の維持がエラーの頻繁な発生により困難となる。そのため、エラー率 20%以上の範囲では、All-D が唯一の ESS 戦略となった。

最後にエラー率が 0 である完全観測の結果と比較する。文献 [10] は割引因子 $\delta = 1$ および完全観測の状況で、2 状態以下の有限オートマトンを戦略集合とした場合に MESS となる戦略は Pavlov-D 戦略とマシン 24 であることを示した*⁹。割引因子が 0 より小さい場合の完全観測下の MESS 結果は既存研究では示されていないが、割引因子 0.95-0.99 の範囲 (0.01 刻み) では Pavlov-D 戦略、マシン 24, All-D 戦略が完全観測下で MESS となることを確認した (All-D 戦略のみは ESS 定義も満たす)。Pavlov-C 戦略は低いエラー率で ESS となるが、エラー率 0 では MESS ともならない結果となった。不完全観測の結果とあわせると、Pavlov-D 戦略と All-D 戦略の MESS はエラー率について 0 付近での連続性を持つ (エラー率 0~ある程度のエラー率まで MESS 定義を満たす) と推測される。

4. Axelrod の進化シミュレーションモデル

4.1 Axelrod モデル

ESS となるマシンの大域的安定性について、Axelrod [6] の行った進化シミュレーションによる検証を行う。ESS は各マシンとのペア対戦の成績により決定され、全プレイヤーが ESS の戦略を採用している状態から他の戦略に侵入されることはない。一方、進化シミュレーションでは、各世代で生き残っているすべてのマシンと対戦した際の平均利得が安定性の基準となる。そのため、各マシンの存在比率も生き残りに影響する。

Axelrod は、繰り返し囚人のジレンマのコンピュータトーナメント (文献 [6], [19]) を開催し、各分野の研究者にプログラム (=戦略) を募った。優勝プログラムは、しっぺ返し戦略 (図 A.1 のマシン 2) であった。この戦略は、最初に協利行動 C を選び、その後は前期の相手の行動を模倣する。Axelrod はしっぺ返し戦略が本当に強い戦略かを検証

するため、高得点の戦略が次のトーナメントで高いシェアを占めるような繰り返し囚人のジレンマトーナメントを繰り返すという進化シミュレーションを行った。1 回 1 回のトーナメントを世代、戦略を生物と見なすと、淘汰の過程ととらえられる。戦略の生き残りやすさをこのような淘汰の過程を利用して検証した。また、新しい振舞いをする行動 (突然変異) はないと仮定した。

世代 τ で戦略 $m \in M$ を選んでいるプレイヤーの比率を $C_m(\tau)$ とする。各世代 τ での存在比率には

$$C_m(0) = \frac{1}{|M|}, \quad \sum_{m \in M} C_m(\tau) = 1 \quad \forall \tau \in \{1, 2, \dots\}$$

という制約を置く。最初の制約は、初期状態では各戦略が等しい比率であることを示す。世代 τ における戦略 m の平均利得 $F_m(\tau)$ は、戦略 m が戦略 k と対戦して得られる利得 $P_1(m, k)$ を用いて

$$F_m(\tau) = \sum_{k \in M} C_k(\tau) P_1(m, k)$$

と表せる。この平均利得を戦略の適応度と見なす。各戦略の再現率 (reproduction rate) は平均利得に比例すると仮定し、 $\tau + 1$ 世代における戦略 $m \in M$ の存在比率を

$$C_m(\tau + 1) = \frac{F_m(\tau) C_m(\tau)}{\sum_{k \in M} F_k(\tau) C_k(\tau)} \quad (3)$$

と定義する。この世代更新式により、高利得の戦略の比率が徐々に高まる戦略淘汰の過程をシミュレーションする。

4.2 分析手法

割引因子 $\delta = 0.99, 0.95$ における各エラー率 0~30% (1%ごと) について、Axelrod モデルのトーナメント 500 世代分のシミュレーションを行った。戦略の利得 $P_1(\cdot, \cdot)$ には、3 章のシミュレーション 1,000 回分で計算された利得の平均値を用いた。戦略は式 (3) に従って各世代で存在比率を進化させる。本論文では、各エラー率での 500 世代時点の戦略の最終存在比率およびエラー率 1%, 10% について戦略の存在比率の推移を示す。以下の節で、各割引因子における結果を考察する。以下の図では、横軸は世代を、縦軸は各世代での戦略の存在比率を表す。

4.3 割引因子 $\delta = 0.99$ における結果と考察

表 3 に割引因子 $\delta = 0.99$ の進化シミュレーションにおいて 500 世代後に観察された主要なマシンの存在比率をエラー率ごとに示す。また、エラー率 1% および 10% の際の存在比率の推移をそれぞれ図 5, 図 6 に示す。

表 3 より、3 章で正のエラー率で ESS と判定されたマシンが進化シミュレーションにおいても生き残りやすい傾向にあることが確認された。また、複数のマシンが ESS と判定されたエラー率でも、ESS マシンの共存はそれほど起こ

*⁹ これは文献 [10] の Figure 4 より確認できる。

表 3 $\delta = 0.99$, 各エラー率 p における 500 世代後の存在比率 (小数第四位四捨五入. エラー率 0.14~0.29 はマシン 4, 12, 26 が正の比率, かつマシン 4, 12 の合計比率が 0.2 以下で推移. 「Pvlv-C」, 「Pvlv-D」は Pavlov-C, Pavlov-D の略記)

Table 3 Relative frequencies after 500 generations for each error rate and $\delta = 0.99$ (Frequencies are rounded off to three decimal places. Machines 4, 12, and 26 have positive frequencies, and the sum of the frequencies of Machines 4 and 12 is at most 0.2 from the error rate 0.14 through 0.29. “Pvlv-C” and “Pvlv-D” denote Pavlov-C and Pavlov-D, respectively).

エラー率 p	マシン番号						
	2 しっぺ返し	3 Pvlv-C	4 トリガー	12	13 All-C	16 Pvlv-D	26 All-D
0	0.152	0.162	0.578	0	0.005	0	0
0.01	0	0.998	0	0	0	0	0
0.02	0	0.999	0	0	0	0.001	0
0.03	0	0.989	0	0	0	0.011	0
0.04	0	0.948	0	0	0	0.052	0
0.05	0	0.981	0	0	0	0.019	0
0.06	0	0.968	0	0	0	0.029	0
0.07	0	0	1.0	0	0	0	0
0.08	0	0	1.0	0	0	0	0
0.09	0	0	1.0	0	0	0	0
0.1	0	0	0.989	0.005	0	0	0.006
0.11	0	0	0.978	0.006	0	0	0.017
0.12	0	0	0.019	0.021	0	0	0.959
0.13	0	0	0.199	0.040	0	0	0.761
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3	0	0	0.028	0.193	0	0	0.777

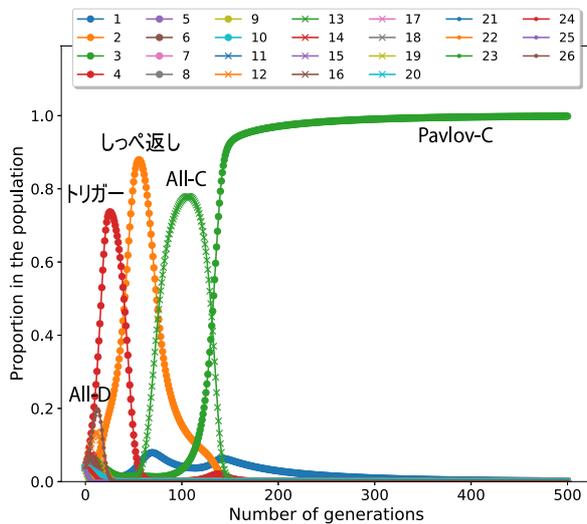


図 5 $\delta = 0.99$, エラー率 1%における存在比率の推移 (横軸: 世代, 縦軸: 存在比率, 図上凡例の数字はマシン番号)

Fig. 5 Evolution of relative frequencies for $\delta = 0.99$ and the error rate 1%. x-axis: generation, y-axis: relative frequency, the numbers in the legend in the upper part of the figure denote the machine numbers.

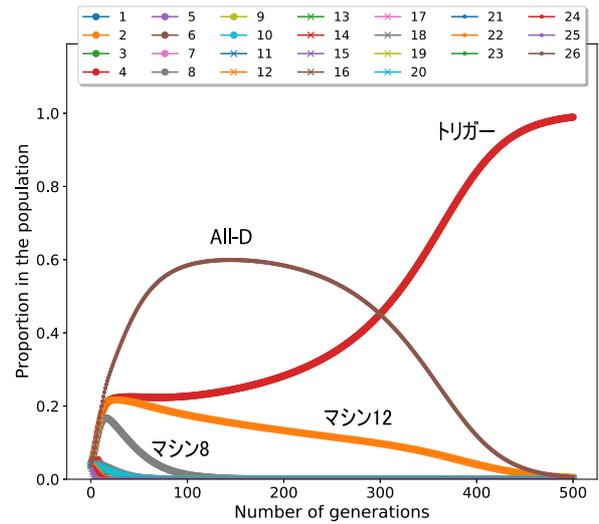


図 6 $\delta = 0.99$, エラー率 10%における存在比率の推移 (横軸: 世代, 縦軸: 存在比率, 図上凡例の数字はマシン番号)

Fig. 6 Evolution of relative frequencies for $\delta = 0.99$ and the error rate 10%. x-axis: generation, y-axis: relative frequency, the numbers in the legend in the upper part of the figure denote the machine numbers.

らず, 1 種類の ESS マシンのみが優勢となる結果となった. 表 3 からは, エラー率 0 と正のエラー率では, エラー率がかなり低い場合でも傾向が大きく異なることが確認できる. エラー率が 0 の場合には, しっぺ返し戦略やトリガー戦略など最初に協調行動をとり, その後も相手が裏切らない限り協調行動を続ける戦略が優勢となった. エラー率 0 において MESS となる Pavlov-D 戦略, マシン 24, All-D 戦略は, しっぺ返しやトリガー戦略に対する利得はそれほど高くなく, 進化シミュレーションでの生き残り戦略とはならなかった. エラー率が正かつ 0.06 以下と比較的低い状況では, Pavlov-C および Pavlov-D 戦略が生き残っており, 判定結果とほぼ一致する. 中程度のエラー率 (0.07~0.11) ではトリガー戦略が生き残りやすい結果となり, これも ESS 判定結果と一致する. 0.12 以上の高いエラー率の範囲では All-D 戦略が優勢となる. また, トリガー戦略と All-D 戦略の生き残りの境界となるのはエラー率 11~12%付近であると考えられる.

図 5 のエラー率 1%の結果では, All-D 戦略とトリガー戦略の存在比率が初期に伸びている. 初期状態では全戦略が等しい比率で分布している. All-D 戦略とトリガー戦略はこの際に得られる総利得が高いため, 初期に存在比率を伸ばす. その後, All-D 戦略とトリガー戦略と対戦した際に多くの利得を得るしっぺ返し戦略が存在比率を伸ばす. その後は, しっぺ返し戦略に対して高い利得を得る All-C 戦略が存在比率を伸ばす. 最終的に, All-C 戦略との対戦で高い利得を得られ, 自身同士の対戦でも高い利得を得られる Pavlov-C 戦略が存在比率を伸ばし, 99%以上のシェアを占める. 3 章の結果より, このパラメータ設定では

Pavlov-C 戦略は ESS である。進化シミュレーションでもこの戦略が出現しやすい戦略であることが判明した。

表 3 より、低エラー率で Pavlov-C 戦略と同様に ESS となる Pavlov-D 戦略は進化シミュレーションでは生き残り戦略とならないか、または低い存在比率にとどまることが分かる。シミュレーション初期に存在比率が伸びない原因は、存在比率が均等の初期状態からは All-D 戦略とトリガー戦略が優勢となる傾向がみられるためである (図 5, 図 6)。裏切り行動から始まる Pavlov-D 戦略はトリガー戦略との相性が悪く、初期に存在比率を伸ばすことができない。また、低エラー率の設定でも中盤以降に優勢とならない原因は、生存戦略が図 5 のように推移するためと考察される。エラー率 0.01~0.06 の状況すべてで、しっぺ返し戦略と All-C 戦略が存在比率を伸ばした後で Pavlov-C 戦略が優勢となる推移が観察された。Pavlov-D 戦略はしっぺ返し戦略と対戦した場合には、エラーが起きない限りは (D, D) の行動プロファイルが継続するため、利得は Pavlov-C 戦略がしっぺ返し戦略と対戦した場合よりも相当に低くなる。この利得差が Pavlov-C 戦略と比較して Pavlov-D 戦略の存在比率が伸びず低い比率にとどまる原因と考えられる。

図 6 のエラー率 10% の結果では、1% の場合と同様に、全戦略と等しく対戦した際の総利得が高いトリガー戦略、マシン 8、マシン 12、All-D 戦略が初期に存在比率を伸ばしている。その後は、この 4 つの戦略どうしで対戦したときに高い利得を得るトリガー戦略と All-D 戦略がさらに存在比率を伸ばす。All-D 戦略は自身と対戦した際には行動プロファイル (D, D) をつねに繰り返し、シグナルエラーがあっても期待利得は 1.3 程度である。一方、トリガー戦略どうしの対戦はシグナルエラーが起こるまで行動プロファイル (C, C) を繰り返すため、All-D 戦略どうしの対戦よりも利得が高くなる。また、All-D 戦略とトリガー戦略どうしの対戦は、早い段階で行動プロファイル (D, D) を繰り返すようになり、2 つの戦略の利得差は少ない。総合してより高い利得を得るトリガー戦略がさらに存在比率を伸ばし、最終的に 99% 以上の比率となる。3 章の結果より、このパラメータでトリガー戦略は ESS であるが、進化シミュレーションでもこの戦略が出現しやすいことが判明した。

エラー率 12% 以上では All-D 戦略が優勢となることが確認された。優勢なマシンがトリガー戦略から All-D 戦略へ切り替わる要因はマシン 8 の動向であると推測される。トリガー戦略、マシン 8、マシン 12、All-D 戦略の初期状態における利得を表 4 に示す。これらのマシンはどのエラー率でも初期状態の利得が他のマシンより高く存在比率を伸ばす傾向にあるが、特にマシン 8 は他のマシンと比較してエラー率が増加するごとの利得の増加が大きい。よってエラー率が高くなるほど、存在比率が初期に伸びやすい。一方で、マシン 8 はシグナル b を観察する限りは C を選び、 g を一度でも観察すると永久に D へ切り替える戦略であ

表 4 $\delta = 0.99$, 各エラー率 p における主要マシンの初期の期待利得 (小数第四位四捨五入)

Table 4 The initial payoff of some selected machines for each error rate and $\delta = 0.99$ (Payoffs are rounded off to three decimal places).

エラー率 p	マシン番号			
	4 (トリガー)	8	12	26 (All-D)
0.1	2.487	2.464	2.490	2.499
0.11	2.485	2.464	2.490	2.500
0.12	2.486	2.466	2.488	2.500
0.13	2.482	2.468	2.491	2.500
0.14	2.484	2.470	2.488	2.500

表 5 $\delta = 0.95$, 各エラー率 p における 500 世代後の存在比率 (小数第四位四捨五入。エラー率 0.08~0.1 はマシン 4 の比率が 1, エラー率 0.13~0.29 はマシン 26 の比率が 1。[Pvlv-C], [Pvlv-D] は Pavlov-C, Pavlov-D の略記)

Table 5 Relative frequencies after 500 generations for each error rate and $\delta = 0.95$ (Frequencies are rounded off to three decimal places. The frequency of Machine 4 is one from the error rate 0.08 through 0.1. The frequency of Machine 26 is one from the error rate 0.13 through 0.29. "Pvlv-C" and "Pvlv-D" denote Pavlov-C and Pavlov-D, respectively).

エラー率 p	マシン番号					
	2 しっぺ返し	3 Pvlv-C	4 トリガー	13 All-C	16 Pvlv-D	26 All-D
0	0.153	0.155	0.590	0	0	0
0.01	0.005	0	0.760	0.235	0	0
0.02	0	0	1.0	0	0	0
0.03	0	0	1.0	0	0	0
0.04	0	0	1.0	0	0	0
0.05	0	0	1.0	0	0	0
0.06	0.143	0	0	0.857	0	0
0.07	0	0	1.0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.11	0	0	1.0	0	0	0
0.12	0	0	0	0	0	1.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3	0	0	0	0	0	1.0

る。そのため、All-D 戦略のマシン 8 に対する利得は比較的高く、初めに C を選ぶトリガー戦略のマシン 8 に対する利得は小さい。高エラー率ではマシン 8 の存在比率が伸びやすく、結果として All-D 戦略に有利な環境となる。これが All-D 戦略が優勢となる要因と推測される。

4.4 割引因子 $\delta = 0.95$ における結果と考察

表 5 に割引因子 $\delta = 0.95$ の進化シミュレーションにおいて 500 世代後に観察された主要なマシンの存在比率をエラー率ごとに示す。また、エラー率 1% および 10% の際の存在比率の推移をそれぞれ図 7, 図 8 に示す。

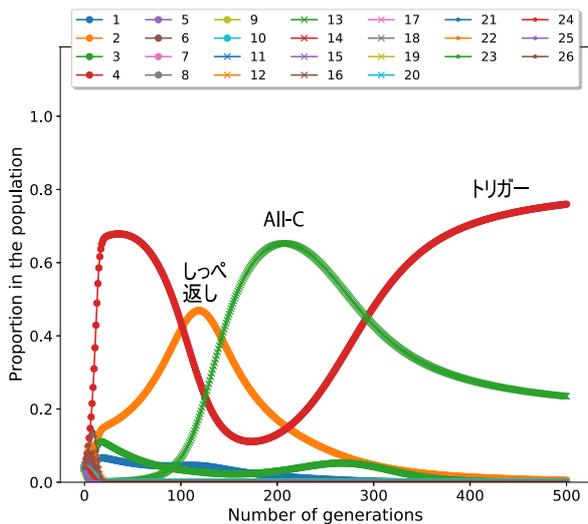


図 7 $\delta = 0.95$, エラー率 1%における存在比率の推移 (横軸: 世代, 縦軸: 存在比率, 図上凡例の数字はマシン番号)

Fig. 7 Evolution of relative frequencies for $\delta = 0.95$ and the error rate 1%. x-axis: generation, y-axis: relative frequency, the numbers in the legend in the upper part of the figure denote the machine numbers.

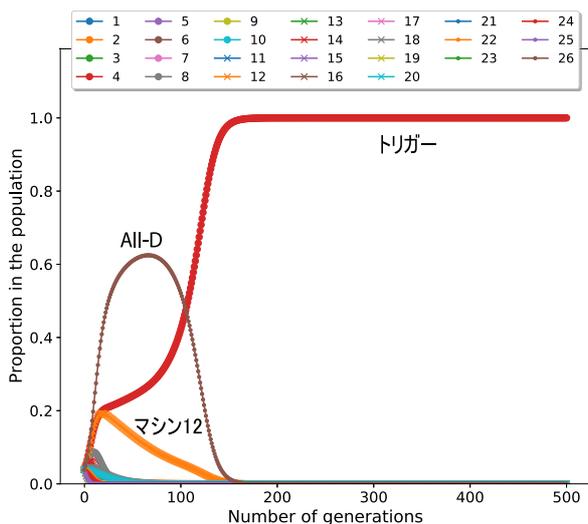


図 8 $\delta = 0.95$, エラー率 10%における存在比率の推移 (横軸: 世代, 縦軸: 存在比率, 図上凡例の数字はマシン番号)

Fig. 8 Evolution of relative frequencies for $\delta = 0.95$ and the error rate 10%. x-axis: generation, y-axis: relative frequency, the numbers in the legend in the upper part of the figure denote the machine numbers.

表 5 より, 多くのエラー率で ESS 判定のマシンが優勢となる, および複数の ESS 判定マシンが存在する場合でも 1 種類が優勢となる場合が多いという $\delta = 0.99$ と同様の特徴がみられる。ただ, $\delta = 0.95$ においては正のエラー率で Pavlov-C 戦略の生き残りが起きず, エラー率が比較的低い状況ではトリガー戦略が, 高いエラー率では All-D 戦略が生き残りやすい傾向がみられた。トリガー戦略と All-D 戦略の生き残りの境界となるのはエラー率 11~12%付近であると考えられる。エラー率が高まると All-D 戦略が優勢と

なる要因は $\delta = 0.99$ の場合と同様である。エラー率 1%では複数の戦略の共存状態が観察された。これは図 7 を用いて詳しく考察する。エラー率 6%については他のエラー率と同様に最終的にトリガー戦略への収束が確認された。これについても 500 世代以降の推移を示し議論する。

図 7 のエラー率 1%における推移では, トリガー戦略が初期に存在比率を伸ばす。その後, トリガー戦略との対戦で高い利得を得るしっぺ返し戦略が比率を伸ばす。さらに, しっぺ返し戦略との対戦で高い利得を得る All-C 戦略が存在比率を伸ばす。最後に, 再びトリガー戦略が存在比率を伸ばす結果となった。図 3 より割引因子 0.95 ではトリガー戦略どうしの利得が比較的高くなるのが分かる。このことが, 最終的なトリガー戦略の生き残りに関連したと考えられる。割引因子が低いほど将来得られる利得が大きく割引かれるため, 相対的に初期の利得が重要となる。たとえば, 割引因子 0.95 の場合は 100 期時点で利得に $(0.95)^{100} \approx 0.006$ が掛けられる。後半の利得割り引きは大きく, 後半の結果は戦略の優劣に影響を及ぼしにくい。トリガー戦略どうしの対戦では初期に行動プロファイル (C, C) を繰り返し比較的高い利得を得やすい。エラー率 1%などシグナルエラーの確率が低い場合には初期の (C, C) が続きやすく, 割引因子が比較的小さいこととあわせてトリガー戦略の利得を高めている。

500 世代時点では, トリガー戦略の存在比率が約 74%, All-C 戦略の存在比率が約 26%となる。この結果は, All-D 戦略に対して多くの利得を得ることができるトリガー戦略が All-D 戦略の存在比率を高くすることを防いでおり, うまくトリガー戦略と All-C 戦略が共存する進化となった。この共存は 500 世代以降も継続し, トリガー戦略・All-C 戦略の存在比率がそれぞれ約 79.4%, 20.6%となる状態で安定して推移する。エラー率 2%以上についてはこのような共存関係はみられず, 下記のエラー率 10%における傾向と同様となった。

図 8 に示すエラー率 10%の推移では, 図 6 に示した割引因子 $\delta = 0.99$ と同様の傾向が観察された。ただし, 存在比率の収束スピードに割引因子の差が表れている。割引因子 0.95 ではトリガー戦略どうしが対戦した際の利得が高い。そのため, 存在比率の増加が 0.99 の場合より速く, 図 6 と比較して図 8 での収束の速さにつながっている。また, このパラメータでトリガー戦略は ESS であるが, 進化シミュレーションでもこの戦略が出現しやすい。

エラー率 6%については, 表 5 からはしっぺ返し戦略と All-C 戦略の共存状態のように読み取れるが, 500 世代以降の進化シミュレーションにより最終的にトリガー戦略への収束が確認された。図 9 に 1,200 世代までの存在比率推移を示す。約 700 世代以降よりトリガー戦略が優勢となっており, これは 1,200 世代以降も継続することを確認した。これより, 11%以下のエラー率については ESS であるトリ

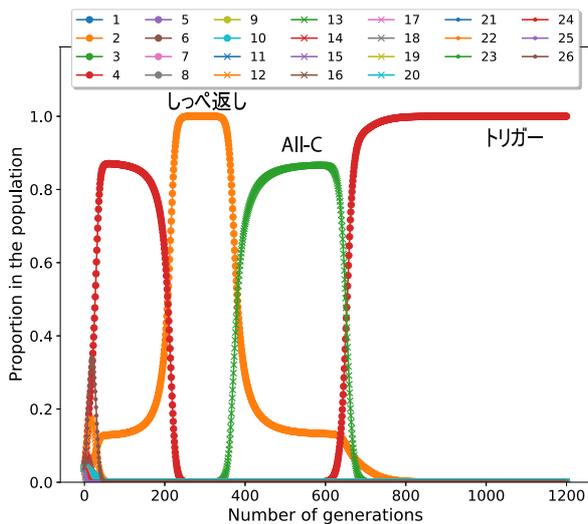


図 9 $\delta = 0.95$, エラー率 6%における存在比率の推移 (横軸: 世代, 縦軸: 存在比率, 図上凡例の数字はマシン番号)

Fig. 9 Evolution of relative frequencies for $\delta = 0.95$ and the error rate 6%. x-axis: generation, y-axis: relative frequency, the numbers in the legend in the upper part of the figure denote the machine numbers.

ガー戦略が進化シミュレーションでも優勢となることが確認できた。

4.5 文献 [20] との結果比較

文献 [20] は, 質量 1 の連続体プレイヤーがお互いに囚人のジレンマの無限繰り返しゲームをプレイする状況における戦略進化を分析している*10. 戦略集合を本研究と同様に 2 状態以下の有限オートマトン戦略に限定している. 連続時間を仮定しており, 戦略の推移は以下の突然変異ありレプリケータダイナミクスで表現される.

$$\dot{C}_m(\tau) = \sum_{k \in M} C_k(\tau) F_k(\tau) q_{km} - C_m(\tau) \sum_{x \in M} C_x(\tau) F_x(\tau)$$

$C_m(\tau)$ は時間 τ での戦略 m の存在比率である. $F_m(\tau)$ は戦略 m の時間 τ における平均利得であり, $F_m(\tau) = \sum_{k \in M} C_k(\tau) P_1(m, k)$ で計算される. q_{km} は戦略 k から m への突然変異率であり, $q_{km} \geq 0$ および $\sum_{m \in M} q_{km} = 1$ である. $\sum_{m \in M} C_m(0) = 1$ であり, $\sum_{m \in M} \dot{C}_m(\tau) = 0$ となるため, 総人口は一定で推移する. 文献 [20] は, 利得構造に応じて以下の 4 つにゲーム状況が分類できることを示した.

- 非協力社会: All-D 戦略が最大多数戦略として出現
 - 不寛容社会: トリガー戦略が最大多数戦略として出現
 - 相互協力社会: Pavlov-C が最大多数戦略として出現
 - 周期協力社会: 複数の戦略が共存または周期的に出現
- 文献 [20] では割引因子 $\delta = 0.9$ を用いており, 本論文の

*10 このような連続体プレイヤーによるゲームは人口ゲームまたは社会ゲームと呼ばれる. これらの簡潔な説明は文献 [35] などが, 詳細な説明は文献 [36] などが参考となる.

設定で値の近い $\delta = 0.95$ の結果を中心に比較をする. 本研究では利得構造は変化させておらず, 仮定した式 (1) の利得は不寛容社会の範囲内となる*11. 表 5 の進化シミュレーションの結果はエラー率中程度まではトリガー戦略が主要な生き残り戦略となっており, この分類と大きな矛盾はない. また, 文献 [20] の図 7 より双方が正しいシグナルを受け取る確率がおおよそ 80% の付近がトリガー戦略が最大多数となる下限であると推測される. 表 5 より, 本研究では $p > 0.11$ であるとトリガー戦略の存在比率が大きく減少する. $p = 0.11$ において双方が正しいシグナルを受け取る確率は約 79% であり, おおむね同様の境界となっている. ただし, 本研究と文献 [20] ではエラーの確率分布の仮定が異なることに注意が必要である. 文献 [20] の主な分析では, いずれかのプレイヤーのみが相手の行動を誤って観測する確率が 1%, 両方のプレイヤーが相手の行動を誤って観測する確率が 3% とされている. また, 双方が正しいシグナルを受け取る確率を増減させる際は, いずれかのプレイヤーのみが相手の行動を誤って観測する確率は 1% と固定し, 両方のプレイヤーが相手の行動を誤って観測する確率を増減して調整している. 本研究では, お互いのシグナルは独立であり, エラー率 1% であれば, 両方のプレイヤーが相手の行動を誤って観測する確率は, $1\% \times 1\% = 0.01\%$ となる. エラー率の仮定が異なるため, 上記議論は推測である. 様々なエラー率分布および割引因子において, 各戦略が優勢となる境界の同定などは今後の検証課題としたい.

また, 文献 [20] の 5.2 節はエラー率を変化させた場合の帰結を分析している. 両プレイヤーが正確に相手の行動を観測できる確率を 0.6 から 0.975 まで変化させ, おおよそ 0.6–0.7 の範囲では非協力社会, 0.725–0.875 の範囲では不寛容社会, 0.9 以上では相互協力社会が出現すると報告されている. シグナル分布の仮定は異なるが, 本研究でも割引因子 0.99 の結果では, エラー率が低い範囲で Pavlov-C 戦略が優勢, 中程度ではトリガー戦略が優勢となる結果を得ており, 優勢となる戦略のエラー率による変化の傾向は整合的である*12.

5. 結論

完全観測下では All-D 戦略が複雑さのコストおよび (1 より小さい割引因子であれば) 割引因子に依存せずつねに理論的に進化的安定となる (文献 [14]). 3 章の結果より, 不完全観測下でも All-D 戦略の進化的安定性に関して同様の強い結果が得られた. しかし, 4 章で検討した均衡から

*11 文献 [20] における $g = l = 0.5$ の利得設定が本研究の設定にあたる.

*12 本研究の割引因子 0.95 の進化シミュレーション結果では, 低いエラー率の範囲でもトリガー戦略が優勢である. 文献 [20] は $\delta = 0.9$ および協利行動の発生しやすい (本論文設定より裏切りの利得, 裏切られたときの損失ともに小さい) 利得設定で上記結果を得ている. 利得設定の違いが, トリガー戦略が優勢となるエラー率の範囲に影響していると推測される.

乖離した状態から最終的に出現する戦略については、全体の戦略分布の時間的遷移も重要となり、ESSのみが決定要因とはならない。進化的安定性による局所的な分析とあわせて、進化シミュレーションなど大域的な安定性の分析の必要性が確認された。

3章と4章の結果より、割引因子が高く、エラー率が小さい場合には、Pavlov-C戦略がESSかつ進化シミュレーションでも高い比率となり、局所的・大局的に見ても最も安定的であった。Pavlov-D戦略は進化シミュレーションでの生き残り戦略とはならず、大域的安定性はPavlov-C戦略に劣るが、より広い範囲のエラー率でESSとなった。エラー率が比較的大きい場合には、トリガー戦略がESSかつ進化シミュレーションでの高い存在比率より、最も安定的といえる。All-D戦略は進化的安定性の点からは非常に強い戦略であり、進化シミュレーションにおいても戦略が均等に分布している初期には存在比率を高めた。しかし、All-D戦略に対抗できる協調的戦略(多くの場合にトリガー戦略)が増えてくると勢いを失う。高いエラー率では生き残り戦略となるが、低いエラー率での大域的な安定性は弱く、他の戦略に淘汰される結果となった。

本研究の課題は、有限オートマトンの状態数の制限緩和である。2状態以下の有限オートマトンは26種類存在し、約340通りの組合せの検討で十分である。3状態以下に緩和すると、マシンの種類は1,322となり、組合せは100万通り近くとなる。現行プログラムでは、計算時間の点で分析不可能であった。この点は本論文の結果を利用した対策が考えられる。たとえば、3章の結果(またあわせて文献[28]の結果)より、Pavlov-C戦略に類似した戦略が高利得でESSとなる可能性が高いことが予想される。特定のマシンのESS判定は、各マシンとの対戦1,322通りの分析で可能である。有望なオートマトンのみの分析とすることで、効率的な判定が可能となる。また、観測エラーを乱数発生させるシミュレーションのため、確定的な判定とならず、シミュレーションごとに誤差が生じた。他アプローチでの利得計算の検討も課題である。

謝辞 加藤滉大氏には分析で使用したプログラムの一部を提供いただいた。また、JSPS 科研費 18K12740 および野村財団(社会科学研究所)の助成を受けている。ご協力に改めて感謝を申し上げます。

参考文献

- [1] Mailath, G.J. and Samuelson, L.: *Repeated Games and Reputations: Long-run Relationships*, Oxford University Press (2006).
- [2] 神取道宏: 人はなぜ協調するのか—繰り返しゲーム理論入門, 三菱研究所 (2015).
- [3] 無藤 望: ゲーム木と戦略的複雑度, オペレーションズ・リサーチ, Vol.58, No.4, pp.198–203 (2013).
- [4] Dal, Bó, P. and Fréchette, G.R.: The evolution of cooperation in infinitely repeated games: Experimental evidence, *American Economic Review*, Vol.101, pp.411–429 (2011).
- [5] Maynard-Smith, J. and Price, G.R.: The Logic of Animal Conflict, *Nature*, Vol.246, pp.15–18 (1973).
- [6] Axelrod, R.: More Effective Choice in the Prisoner's Dilemma, *Journal of Conflict Resolution*, Vol.24, pp.379–403 (1980).
- [7] Andreozzi, L.: Evolutionary stability in repeated extensive games played by finite automata, *Games and Economic Behavior*, Vol.79, pp.67–74 (2013).
- [8] Rubinstein, A.: Finite automata play the repeated prisoner's dilemma, *Journal of Economic Theory*, Vol.39, pp.83–96 (1986).
- [9] Abreu, D. and Rubinstein, A.: The structure of Nash equilibrium in repeated games with finite automata, *Econometrica*, Vol.56, pp.1259–1282 (1988).
- [10] Binmore, K. and Samuelson, L.: Evolutionary stability in repeated games played by finite automata, *Journal of Economic Theory*, Vol.57, pp.278–305 (1992).
- [11] Volij, O.: In defense of defect, *Games and Economic Behavior*, Vol.39, pp.309–321 (2002).
- [12] Cooper, D.J.: Supergames Played by Finite Automata with Finite Costs of Complexity in an Evolutionary Setting, *Journal of Economic Theory*, Vol.68, pp.266–275 (1996).
- [13] García, J. and van Veelen, M.: In and out of equilibrium I: Evolution of strategies in repeated games with discounting, *Journal of Economic Theory*, Vol.161, pp.161–189 (2016).
- [14] van Veelen, M. and García, J.: In and out of equilibrium II: Evolution in repeated games with discounting and complexity costs, *Games and Economic Behavior*, Vol.115, pp.113–130 (2019).
- [15] Piccione, M. and Rubinstein, A.: Finite Automata play a repeated extensive game, *Journal of Economic Theory*, Vol.61, pp.160–168 (1993).
- [16] Muto, N.: Strategic complexity in repeated extensive games, *Games and Economic Behavior*, Vol.83, pp.45–52 (2014).
- [17] Nowak, M.A., Sigmund, K. and El-Sedy, E.: Automata, repeated games and noise, *Journal of Mathematical Biology*, Vol.33, pp.703–722 (1995).
- [18] Posch, M.: Win-Stay, Lose-Shift Strategies for Repeated Games—Memory Length, Aspiration Levels and Noise, *Journal of Theoretical Biology*, Vol.198, No.2, pp.183–195 (1999).
- [19] Axelrod, R.: Effective Choice in the Prisoner's Dilemma, *Journal of Conflict Resolution*, Vol.24, pp.3–25 (1980).
- [20] 西野上和真, 五十嵐瞭平, 岩崎 敦: 私的観測下の繰り返し囚人のジレンマにおける協力のダイナミクス, 情報科学技術フォーラム講演論文集 (FIT2020), Vol.19, No.2, pp.27–34 (2020).
- [21] 四辻 隆, 澤 亮治: 不完全観測下における損失回避性が均衡の安定性に及ぼす影響, 情報科学技術フォーラム講演論文集 (FIT2021), Vol.20, No.2, pp.65–72 (2021).
- [22] 五十嵐瞭平, 岩崎 敦: ほほ公的観測下の囚人のジレンマにおける協力のダイナミクス, 情報科学技術フォーラム講演論文集 (FIT2021), Vol.20, No.2, pp.81–88 (2021).
- [23] 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).
- [24] Boerlijst, M.C., Nowak, M.A. and Sigmund, K.: The Logic of Contrition, *Journal of Theoretical Biology*, Vol.185, No.3, pp.281–293 (1997).
- [25] Imhof, L.A., Fudenberg, D. and Nowak, M.A.: Tit-for-

tat or win-stay, lose-shift?, *Journal of Theoretical Biology*, Vol.247, No.3, pp.574–580 (2007).

[26] Zagorsky, B.M., Reiter, J.G., Chatterjee, K. and Nowak, M.A.: Forgiver Triumphs in Alternating Prisoner’s Dilemma, *PLOS ONE*, Vol.8, No.12, e80814 (2013).

[27] Kandori, M. and Obara, I.: Towards a Belief-Based Theory of Repeated Games with Private Monitoring: An Application of POMDP, *Working Paper*, University of California Los-Angeles (2010).

[28] ジョンジョン, 岩崎 敦, 神取道宏, 小原一郎, 横尾 真: 部分観測可能マルコフ決定過程を用いた私的観測付き繰返しゲームにおける均衡分析プログラム, *情報処理学会論文誌*, Vol.53, No.11, pp.2445–2456 (2012).

[29] Compte, O.: Communication in repeated games with imperfect private monitoring, *Econometrica*, Vol.66, pp.597–626 (1998).

[30] Kandori, M. and Matsushima, H.: Private Observation, Communication and Collusion, *Econometrica*, Vol.66, pp.627–652 (1998).

[31] Piccione, M.: The repeated prisoner’s dilemma with imperfect private monitoring, *Journal of Economic Theory*, Vol.102, pp.70–83 (2002).

[32] Ely, J.C. and Valimaki, J.: A robust folk theorem for the prisoner’s dilemma, *Journal of Economic Theory*, Vol.102, pp.84–105 (2002).

[33] Sugaya, T.: Folk Theorem in Repeated Games with Private Monitoring, *Review of Economic Studies* (公刊予定) (2021).

[34] Kraines, D. and Kraines, V.: Pavlov and the prisoner’s dilemma, *Theory and Decision*, Vol.26, pp.47–79 (1989).

[35] 尾山大輔: 社会ゲームにおけるナッシュ均衡の安定性: ポテンシャル・ゲームと完全予見動学, *計測と制御*, Vol.55, No.4, pp.362–367 (2016).

[36] Sandholm, W.H.: *Population Games and Evolutionary Dynamics*, MIT Press (2010).

付 録

A.1 2 状態以下の有限オートマトン

2 状態以下の有限オートマトンを図 A.1 に示す. 論文内の番号付きのマシン名は図内の番号に対応している.

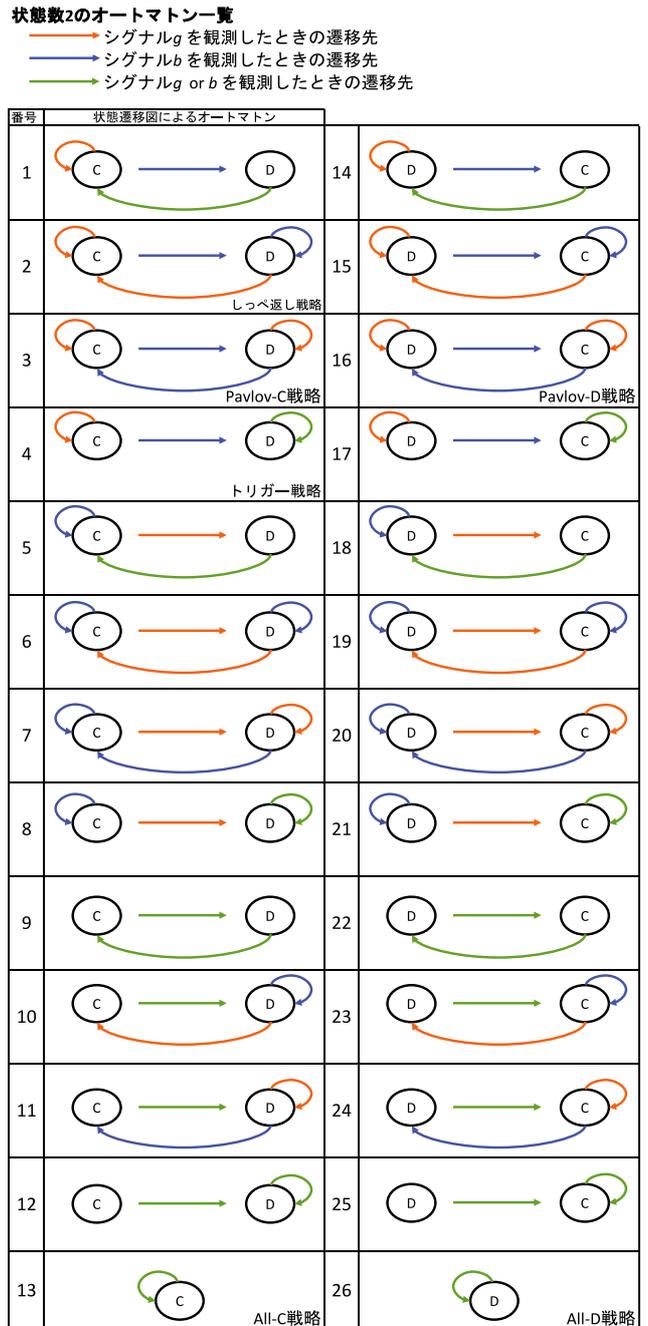


図 A.1 2 状態以下の有限オートマトン

Fig. A.1 Finite automata with at most two states.



小池 淳平 (正会員)

2019 年福島大学理工学群共生システム理工学類卒業. 2021 年筑波大学システム情報工学研究科社会工学学位プログラム博士前期課程修了. 在学中, ゲーム理論に関する研究に従事.



澤 亮治 (正会員)

1998年慶應義塾大学工学部電気工学科卒業。2012年Wisconsin州立大学Madison校経済学博士課程修了，経済学博士。会津大学コンピュータ工学部文化研究センター准教授を経て，2016年より筑波大学システム情報系社会工学域准教授。ゲーム理論，社会学習理論に関する研究に従事。