CF-002

見間違えのあるくり返し囚人のジレンマにおける確率動学にもとづく戦略進化

谷川 颯希 *
Satsuki Tanikawa

岩崎 敦 *
Atsushi Iwasaki

概要

本論文は、見間違えのあるくり返し囚人のジレンマにおける戦略の進化を確率動学にもとづいて吟味する。見間違えは，プレイヤが相手の行動についてノイズを含むシグナルを観測し，そのシグナルを他のプレイヤは観測できないという特徴をもつ．ここで, どんな戦略の組が均衡になるかはゲーム理論の有名な未解決問題の一つであり，本研究では戦略空間を状態数 3 以下の有限状態機械に限定した確率動学の帰結から、どのような戦略が生き残るかを吟味した．その結果，見間違えのない場合は、常に裏切り（ALLD）や一度でも裏切られたら許さない（Grim-trigger，GRIM）といった戦略が最大多数を占めるが、他の戦略も一定のシェアをもち、特定の戦略が集団を支配することはなかった。一方で、見間違えのある場合は、利得構造に応じて集団を支配する戦略が現れるようになった。とくに状態数 2 以下の戦略空間では、裏切られたと感じたら、1 回裏切ってから協力に戻る 1 期相互処罰戦略が、状態数 3 以下の場合では、その拡張である 2 回連続で裏切ってから協力に戻る 2 期相互処罰戦略が、広範囲のパラメータに渡って集団を支配することがわかった。また、これらの戦略への段階的な進化の様子を明らかにした。

1　はじめに

無限回くり返しゲームは，長期的関係にあるプレイヤ間の（暗黙の）協調を説明するためのモデルである [12]．主に経済学分野で企業間の談合といった協調行動を分析するために発展してきた [15]。暗黙の協調を実現するには、プレイヤが相手の行動をある程度観測できることが前提となる。これまで、見間違えのない、つまり相手の行動が完全に観測できる完全観測（perfect monitoring）のケースについては多く論じられている [1, 7, 9]。しかし、現実には相手の行動が完全に観測できない不完全観測（imperfect monitoring）のケース、つまり、プレイヤが相手の行動についてノイズを含むシグナルを観測し、そのシグナルを他のプレイヤは観測できない場合がある。これはとくに、見間違えのない、不完全私的観測（imperfect private monitoring）のケースと呼ばれる [4, 13, 16]。このゲーム（infinite repeated games with imperfect private monitoring）の特徴は、プレイヤが相手の行動に関してノイズを含む観測（シグナル）を私的に受け取ると仮定する点にある。いいかえると、あるプレイヤが相手の行動について観測したシグナルと異なるシグナルを他のプレイヤが観測しているかもしれない。不完全私的観測付き無限回くり返しゲームおいてどのような振る舞い (戦略) が均衡になるのかについては、ゲーム理論における代表的なゲームである囚人のジレンマの例でさえ十分にわかっていない。例えば、部分観測可能マルコフ決定過程 (Partially Observable Markov Decision Process, POMDP) を用いて均衡を計算する手法 [12] が知られているが、その計算量は一般には決定不能と知られている。

本論文では、均衡の代わりに理論生物学や進化ゲーム理論で広く用いられている確率動学 (stochastic dynamics) [14, 8] にもとづいて、見間違えのあるくり返し囚人のジレンマにおいて、どのような戦略が長期的に生き残るかを分析する。確率動学は、有限集団における戦略頻度の確率的な時間変化を記述するモデルであり、個体が利得に応じて戦略を模倣・更新するプロセスを通じて進化が進行する。このような進化は、マルコフ連鎖によってモデル化されることが多く、長期的には定常分布や吸収状態が観察される。このアプローチにより、進化的に安定な戦略だけでなく、戦略の持続的な出現確率や集団構成における支配度を分析できる。また、均衡分析と異なり、戦略が多様に入れ替わる環境でも、淘汰圧のかかり方に応じてどのような戦略が実効的に優勢となるかを評価できる。本研究では、こうした確率動学の枠組みに基づき、見間違えのある環境において、どのような戦略が進化的に優位となり、最終的に集団を支配するかを明らかにする。

加えて、確率動学は均衡分析とは異なる視点を提供する。厳密には、均衡と動学的帰結のあいだに包含関係は存在しない。すなわち、ある戦略の組がナッシュ均衡を構成していたとしても、それが実際の動学の過程において最大多数を占めるとは限らず、その逆もまた真である。さらに、均衡が複数存在する場合には、どの均衡にダイナミクスが収束するかを事前に予測することは困難であり、場合によっては均衡を構成する戦略自体が動学的には現れないこともある。このような状況において、自然淘汰の力学（淘汰圧）に基づく確率動学は、どの戦略が環境内で生き残りやすいか、またどのような戦略が実効的に集団内で支配的となるかを示す有力な手段である。

理論生物学や進化ゲームの文脈では、行動の取り違え、つま

* 電気通信大学, The University of Electro-Communications

り自分の行動を出し間違える摂動 (trembling-hand) は盛んに研究されてきた [9, 2, 3]。しかし、その重要性にも関わらず、相手の行動を見間違える私的観測を進化ゲームの文脈で網羅的に分析することは非常に難しいと考えられてきた。その理由の 1 つとして、摂動と私的観測は戦略と情報の構造が異なるため、従来の成果が適用できないことに挙げられる。また、一般には複雑な行動計画となるくり返しゲームの戦略を有限状態機械 (Finite State Automaton, FSA) で記述するとき、私的観測をどのようにモデル化し期待利得を計算するかよくわかっていなかった。そこで本論文では、まずプレイヤが取りうる戦略を状態数 3 以下の FSA に限定する。つまり、プレイヤの今日とった行動と観測したシグナルから明日の行動への写像を考える。戦略を FSA に限定したときの期待利得をマルコフ決定過程に基づいて計算し、その利得表をもとに確率動学を計算する。

その結果、見間違えのない場合は、常に裏切り（ALLD）や一度でも裏切られたら許さない（Grim-trigger, GRIM）といった戦略が最大多数を占めるが、他の戦略も一定のシェアをもち、特定の戦略が集団を支配することはなかった。一方で、見間違えのある場合は、利得構造に応じて集団を支配する戦略が現れるようになった。とくに状態数 2 以下の戦略空間では、裏切られたと感じたら、1 回裏切って協力に戻る 1 期相互処罰戦略 (1-period Mutual Punishment, 1MP) [12] が、状態数 3 以下の場合では、その拡張である 2 回連続で裏切ってから協力に戻る 2 期相互処罰戦略 (2MP, #734) が、広範囲のパラメータに渡って集団を支配することがわかった。興味深いことに、割引因子を 1 未満に固定した本研究のモデルでは、これらの戦略は裏切り開始する戦略 (Suspicious-1MP, -2MP) が生き残りやすいことがわかった。これは、最初に裏切る戦略と対戦したとき、自分も最初に裏切るようにすることで、裏切られることにより損失を防ぐためである。

さらに、これらの戦略への段階的な進化を観察したところ、状態数 2 以下の戦略空間では、ALLD から GRIM、GRIM から TFT、TFT から ALLC もしくは FGV、それから S-1MP へと段階的に進化した。ここで FGV とは Forgiver と呼ばれ、裏切られたと感じたら、1 回裏切ってから自動的に協力に戻る ALLC より厳しく、1MP や TFT より寛容な戦略である。

状態数 3 以下の戦略空間では、1054 個もの戦略が存在するため、その段階的進化を観察するのは困難である。そこで、高次元データを 2 次元に次元削減アルゴリズムの 1 つである t-SNE(t-Distributed Stochastic Neighbor Embedding) を用いて、固定確率 (fixation probability) の高次元ベクトルを 2 次元に落とし込んだ。ここで固定確率とは、ある突然変異型（または新たな戦略）が導入されたときに、その型が将来的に集団全体を占める（すなわち、他の型を駆逐して定着する）確率であり、本論文での確率動学の帰結はこの固定確率行列の定常分布で表される。これにより 1054 戦略の空間における戦略の段階的進化をある程度観察することに成功した。その結果、ALLD 型の戦略が裏切りから始まる TFT (Suspicious TFT, S-TFT) 型に遷移した後、ALLC 型に遷移し、S-2MP の周辺に収束する。さらにノイズなどに関する感度分析から十分広いパラメータに渡って同じ傾向を保つことがわかった.

表 1: 囚人のジレンマ $(g > 0,\ l > 0\ and\ |g - l| < 1)$

	$a_2 = C$	$a_2 = D$
$a_1 = C$	$1,\ 1$	$-l,\ 1 + g$
$a_1 = D$	$1 + g,\ -l$	$0,\ 0$

表 2: (C, C) のときのシグナル分布

	$w_2 = g$	$w_2 = b$
$w_1 = g$	$(1-\varepsilon)^2$	$(1-\varepsilon)\varepsilon$
$w_1 = b$	$\varepsilon(1-\varepsilon)$	ε^2

2 モデル

本章では文献 [12] に基づいて、見間違えのある無限回くり返しゲームをモデル化する。ここでプレイヤ $k \in \{1, 2\}$ は成分ゲームを無限期間 $t = 0, 1, 2, \cdots$ に渡って繰り返す。各期においてプレイヤ k は有限集合 $A = \{C, D\}$ から行動 a_k を選択し、その行動の組を $\boldsymbol{a} = (a_1, a_2) \in A^2$ とする。次に、プレイヤ k は $\boldsymbol{a}$ に関する私的なシグナル $\omega_k \in \Omega$ を観測する。$\boldsymbol{w}$ をシグナルの組 $(\omega_1, \omega_2) \in \Omega^2$ とする。また、プレイヤが $\boldsymbol{a}$ を選択したとき $\boldsymbol{w}$ が生起する同時確率を $o\left(\boldsymbol{w} \mid \boldsymbol{a}\right)$ とし、この同時確率を与える分布のことをシグナル分布と呼ぶ。成分ゲームは無限回くり返し行われるので、プレイヤ k の割引利得和は割引因子 $\delta \in (0, 1)$ により $\sum_{t=1}^{\infty} \delta^t g_k(\boldsymbol{a}^t)$ となる。$g_k(\cdot)$ は表 1 に示す囚人のジレンマの利得表に従う。

次にプレイヤ 2 の行動に関するプレイヤ 1 のノイズを含む観測をプレイヤ 1 の私的シグナルとし、$\omega \in \{g, b\}$ $(good, bad)$ とする。正しい観測ではプレイヤ 2 が C を選択した際のプレイヤ 1 の私的シグナルは g, D を選択した際の私的シグナルは b となる。プレイヤ 2 についても同様である。不完全私的観測の分野においてよく使われる分布に、条件付き独立観測がある [5]。ここでは、各プレイヤが行動を見間違える確率を ε とする。例として、(C, C) が実現した場合のシグナル分布を表 2 に示す。

プレイヤの戦略は、そのプレイヤの過去の行動と受け取ったシグナルから現在の行動への写像で表現される。無限回くり返しゲームの戦略は、文字通り無限個存在し、その全てを網羅するのは不可能である。このため、先行研究ではプレイヤの戦略を何らかの形で制限している [11, 6]。そこで本論文では、有限状態機械 (Finite State Automaton, FSA) による戦略表記を採用し、確率動学で計算可能な戦略空間を定義する。同相な FSA とはまったく同じ行動パターンをとる FSA のことを言い、同相な FSA も含めて列挙すると戦略の個数は、Θ を FSA の状態数として、$|A||\Theta|^{|A|}|\Theta|^{|\Theta||\Omega|}$ となる。これに対して、同相な FSA をまとめると状態数が 2 の場合は 26 個、状態数が 3 の場合は 1054 個の非同相な FSA が戦略空間を定義する。

図 1 に 3 状態以下の FSA 戦略のうち、主要なものを示す。図 1a に示す ALLD は常に裏切る戦略であり、図 1f に示す ALLC は常に協力する戦略である。図 1b に示す GRIM は最

初はシグナル g を観測する限り協力するが、シグナル b を観測するとその後は永遠に裏切り続ける戦略である。

図 1c はに示す S-1MP は最初は裏切り、シグナル g を観測すると自分が直前に選択した行動と同じ行動をとり、シグナル b を観測すると自分が直前に選択した行動とは異なる行動をとる戦略である。図 1d に示す S-TFT は最初は裏切り、シグナル g を観測すると協力し、シグナル b を観測すると裏切る戦略である。図 1e に示す S-FGV は最初は裏切り、その後はシグナル g を観測する限り協力するが、シグナル b を観測すると一度裏切って再び協力に戻り、シグナル g を観測する限り協力する戦略である。

次に、3 状態以下の FSA 戦略空間において重要な戦略を紹介する。図 2a に示す#734 は状態 R ではシグナル g を観測する限り協力するが、シグナル b を観測すると、その後シグナル b を二回連続で観測するまでは裏切り続け、二回連続でシグナル b を観測すると状態 R に戻る戦略である。図 2b に示す#736 は状態 R ではシグナル g を観測する限り協力し、シグナル b を観測すると次はシグナル g を観測する限り裏切り続ける。そして、再びシグナル b を観測するともう一度裏切って状態 R に戻る戦略である。#734 は状態 R でシグナル b を観測すると、その後はシグナル b を二回連続で観測するまではずっと裏切り続けるが、#736 は状態 R でシグナル b を観測すると、その後はシグナル b 再び観測すると一度裏切ってすぐに状態 R に戻るという違いがある。

数ある戦略の中から有効な戦略を発見する方法の一つとして、確率動学がある。文献 [14] にもとづいて有限集団における戦略の帰結を吟味するための確率動学を概説する。この戦略分布の動学は出生死亡過程で表現され、毎期に 1 個体が死亡し、1 個体が新たに生まれるとする。死亡する個体は有限集団の中からランダムに決まる一方で、生まれてくる個体がもつ戦略はその戦略がその集団で獲得する利得に応じた確率で決まる。ここで、各個体は 26 個 (1054 個) の戦略のうち 1 つをもち、その利得は集団内で総当りでゲームを無限回プレイした結果（戦略同士の割引利得和）から計算する。この出生死亡過程はエルゴード性をもつことが知られており、ある戦略の集団 1 個体だけ異なる戦略を侵入させたとき、その戦略がもともといた戦略を淘汰する確率（固定確率）からなる確率行列を構成することで, その帰結を計算できる。つまり、任意の 2 つの戦略 m および n の間の固定確率 $\rho_{m,n}$ を、戦略 m をもつプレイヤの集団が全て戦略 n をもつプレイヤに侵略される確率とする。出生死亡過程の帰結となる戦略分布は、戦略の数を S とすると、以下の要素をもつ行列 T の転置行列 T' の固有ベクトルを計算することで求められる：

$$T_{mn} = \begin{cases} \frac{1}{S-1}\,\rho_{m,n} & (m \neq n) \\ 1 - \frac{1}{S-1}\sum_{m\neq n}\rho_{m,n} & (m = n) \end{cases}.$$

3　見間違えのない環境下での確率動学

本章では、プレイヤが相手の行動を完全に観測できる環境（見間違えが発生しない場合, $\varepsilon = 0.0$）において、FSA で表現される戦略の進化的な帰結を確率動学の枠組みで分析した。戦略空間は状態数 2 および 3 以下の FSA 戦略に限定し、各戦略間の固定確率をもとにマルコフ過程を構成し、長期的な定常分布を計算することで、集団内における戦略の分布と協力行動の出現傾向を明らかにした。

状態数 2 以下の戦略空間においては、利得構造を表す、裏切りによって得られる利得の増分（gain, g）や裏切られることによる損失（loss, l）の大きさに応じて、ALLD や GRIM が最大多数戦略として現れる傾向が確認された（図 3a)。たとえば $g = l = 1.00$ のとき、ALLD の戦略比率は 57% に達するが、協力率は 0.40 にとどまるなど、非協力的な戦略が優勢となる状況が観察された。一方で、g, l が中程度では、GRIM に加えて 1MP、FGV、TFT などが一定比率で共存し、協力率が最大 0.91 に達するような高協力状態も出現した。

状態数 3 以下の戦略空間に拡張した場合、代表的な 2 状態戦略が最大多数になることはないが、S-2MP (#734) や#736 といった戦略を中心に、一定の条件下で複雑な罰行動を含む戦略が最大多数となる傾向が観察された。たとえば、特定のパターンの観測が揃うまで裏切りを継続し、その後協力に戻るような戦略群が一定比率で共存していた。利得構造に依存しつつも協力率はおおむね 0.70〜0.85 の範囲で推移しており、特に複雑な罰行動を含む戦略の出現により、高い協力水準が保たれる傾向が確認された。

これらの結果は、一見すると多様な戦略が共存しているように見えるが、見間違えが発生しない環境では、それらの多くが「相手が協力する限り自らも協力を続ける」といった類似の振る舞いを示しており、実際に実現される行動パターンには多様性が実は乏しい。すなわち、戦略の記述上の多様性にもかかわらず、進化的に安定となる行動様式は限定的である。

4　見間違えのある環境下での確率動学の帰結

4.1　2 状態

図 4 に行動の見間違えがあるときの戦略空間を 2 状態以下の FSA 戦略に限定した行動を見間違えうる場合の確率動学の帰結を示す。g および l を $[0.00, 1.00]$ で、$|g - l| < 1$ を満たしつつ、0.01 刻みで変化させた。また、プレイヤが行動を見間違える確率を $\varepsilon = 0.05$ に固定した。最大多数戦略を図 4a に、協力率を図 4b に示す。図 4c-4g には図 4a において最大多数戦略となった戦略および主要な戦略の戦略比率を、図 4h にはその他の戦略の比率の合計を示している。まず、g および l が十分に大きいとき、ALLD が最大多数戦略となる。$g = l = 0.80$ のとき、ALLD の戦略比率は 0.86 となり、協力率は 0.08 となる。このとき、裏切る誘因や裏切られることによる損失が大きいため、他のどの戦略も協力を維持するに十分な将来利得を獲得できない。g が大きく、l が小さいとき、GRIM が最大多数

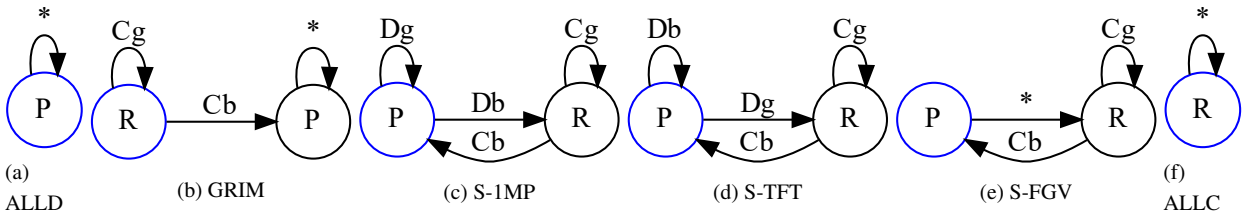


図 1: 状態数 2 以下の主要な FSA 戦略

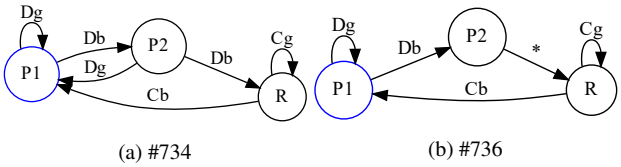


図 2: 状態数 3 の主要な FSA 戦略

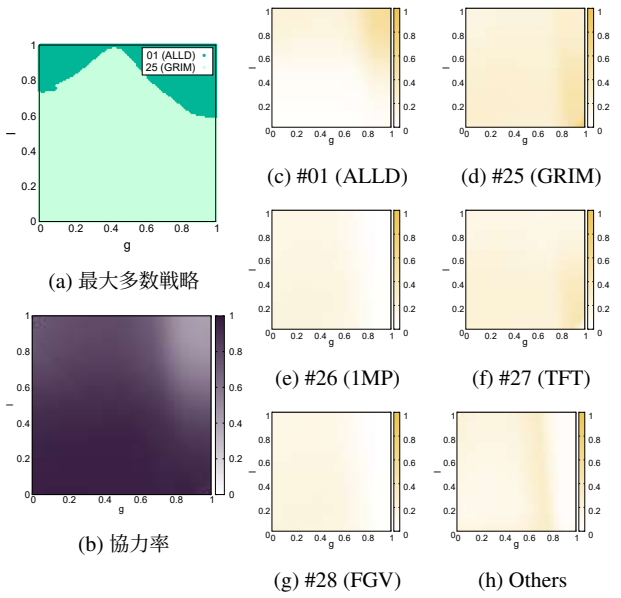


図 3: 戦略空間を 2 状態以下の FSA 戦略に限定した場合の行動の見間違えが起こらない場合の確率動学の帰結 ($N = 100, s = 10, \delta = 0.90, \text{and } \varepsilon = 0.00$)

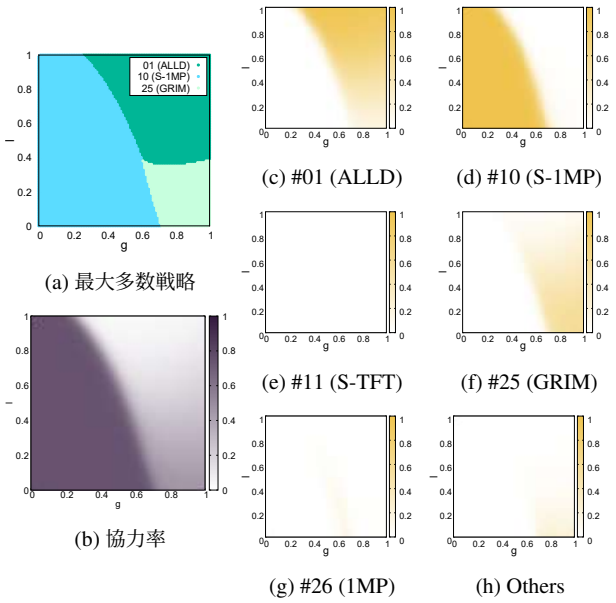
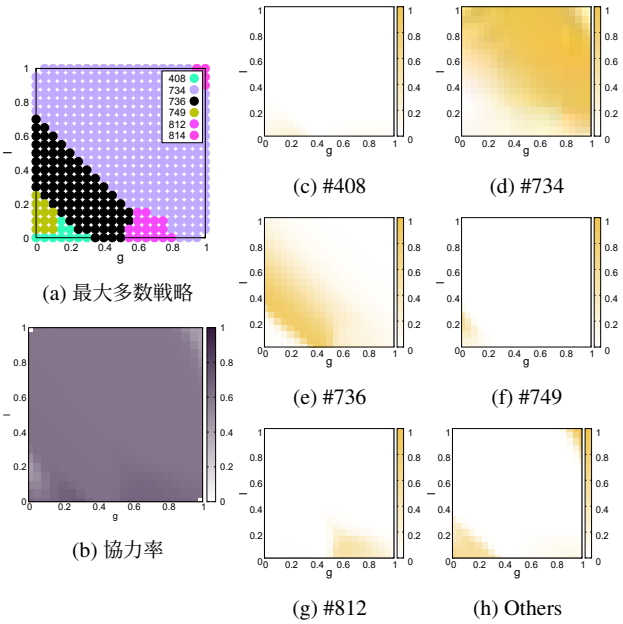


図 4: 戦略空間を 2 状態以下の FSA 戦略に限定した場合の行動を見間違えうる場合の確率動学の帰結 ($N = 100, s = 10, \delta = 0.90, \text{and } \varepsilon = 0.05$)

戦略となる。$g = 0.80, l = 0.20$ のとき、GRIM の戦略比率は 0.53 となり、協力率は 0.38 となる。裏切られることによる損失が小さくなったことで、最初は協力するが相手の裏切りを観測するとその後は永遠に裏切り続ける戦略である GRIM が生き残るようになる。そして、それ以外の広い範囲では S-1MP が最大多数戦略となる。$g = l = 0.2$ のとき、S-1MP の戦略比率は 1 であり、協力率は 0.74 となる。S-1MP 同士の対戦では見間違えが起こらなければ、相互協力状態を維持することができ、見間違えの発生によって相互協力状態が途切れたとしても、お互いに一度裏切り合うことで、相互協力状態を回復することができる。そのため、S-1MP は高い戦略比率を獲得すると考えられる。

4.2 3 状態

図 5 に戦略空間を 3 状態以下の FSA 戦略に限定した行動を見間違えうる場合の確率動学の帰結を示す。g および l を $[0.00, 1.00]$ で、$|g - l| < 1$ を満たしつつ、0.01 刻みで変化させた。また、プレイヤが行動を見間違える確率を $\varepsilon = 0.05$ に固定した。最大多数戦略を図 5a に、協力率を図 5b に示す。図 5c-5g には図 5a において最大多数戦略となった戦略の戦略比率を、図 5h にはその他の戦略の比率の合計を示している。図 5a から分かるように戦略空間を 2 状態以下の FSA 戦略に限定した場合に最大多数戦略となる戦略は戦略空間を 3 状態以下の FSA 戦略に限定した場合は最大多数戦略とはならない。g および l が小さいとき、#408 や#749 が最大多数戦略となる。$g = l = 0$ のとき、#408 の戦略比率は 0.13 となり、協力率は 0.66 となる。g および l が中程度になると、#736 や #812 が最大多数戦略となる。$g = l = 0.2$ のとき、#736 の戦略比率は 0.85 となり、協力率は 0.62 となる。g および l がさ



(a) 最大多数戦略　(c) #408　(d) #734

(b) 協力率　(e) #736　(f) #749

(g) #812　(h) Others

図 5: 戦略空間を 3 状態以下の FSA 戦略に限定した場合の行動を見間違えうる場合の確率動学の帰結 ($N = 100, s = 10, \delta = 0.90,$ and $\varepsilon = 0.05$)

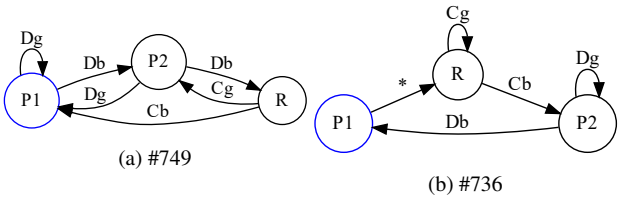


図 6: 見間違えのある環境下で生き残る戦略

らに大きくなると、広い範囲で#734 が最大多数戦略となる。$g = l = 0.8$ のとき、#734 の戦略比率はほぼ 1 となり、協力率は 0.59 となる。

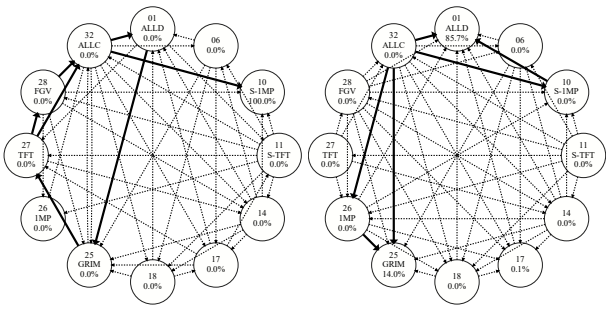
5　見間違えのある環境下での戦略の進化

5.1　2 状態

図 7 に戦略空間を 2 状態以下の FSA 戦略に限定した行動を見間違えうる場合の結果を示す。ゲインとロスが小さいとき、図 4 で見たように、S-1MP が支配的になる。図 7a では、S-1MP が他の戦略を直接侵略するようにはなっていないが、ALLD を出発点としたとき、GRIM, TFT, FGV, ALLC などを経由して S-1MP に到達している。ゲインとロスが大きくなると図 4 で見たように ALLD, GRIM の混合が支配的になる。図 7b において、ALLC が様々な戦略に侵略されたのち、1MP は GRIM に、S-1MP は ALLD に侵略されるようになる。

5.2　3 状態

本節では、3 状態以下の FSA 戦略がどのように淘汰され進化していくかを分析する。特に、戦略間の淘汰関係を視覚



(a) $g = 0.2$ and $l = 0.2$　(b) $g = 0.8$ and $l = 0.8$

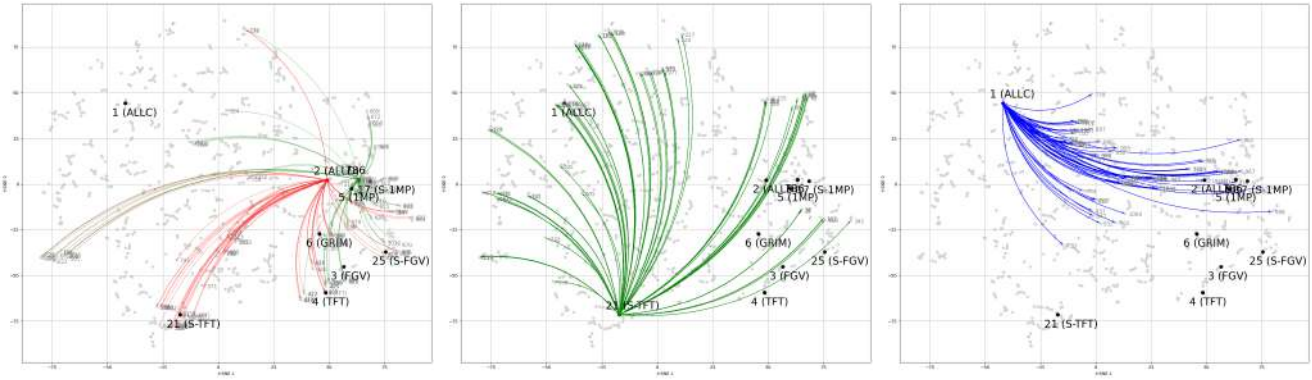
図 7: 行動を見間違えうる場合の 2 状態以下の FSA 戦略の進化 ($N = 100, s = 10, \delta = 0.9,$ and $\varepsilon = 0.00$)

的に把握するため、1054 個の戦略に対する固定確率ベクトル（1054 次元）を、次元削減手法である t-SNE（t-distributed Stochastic Neighbor Embedding）[10] により 2 次元に写像し、その空間上に淘汰関係を描画している（図 8, 図 9）。各ノードは 1 つの戦略を表し、ある戦略から他の戦略へ固定確率が高い順に上位 64 個を選び、淘汰される向きにアーク（有向辺）を引いている。

図 8(a) は、ALLD（常に裏切る戦略）と、S-2MP 型の代表戦略である#734 および#736 の淘汰経路を示している。興味深いのは、固定確率ベクトル空間においては ALLD と#734, #736 は近接しており、いずれも似た淘汰圧を受ける立場にあることが t-SNE 上で確認される点である。しかし、淘汰される側としての挙動は大きく異なる。ALLD は S-TFT やその周辺に位置する戦略によって淘汰されやすく、進化の過程において短期間でその比率を失う傾向が強い。

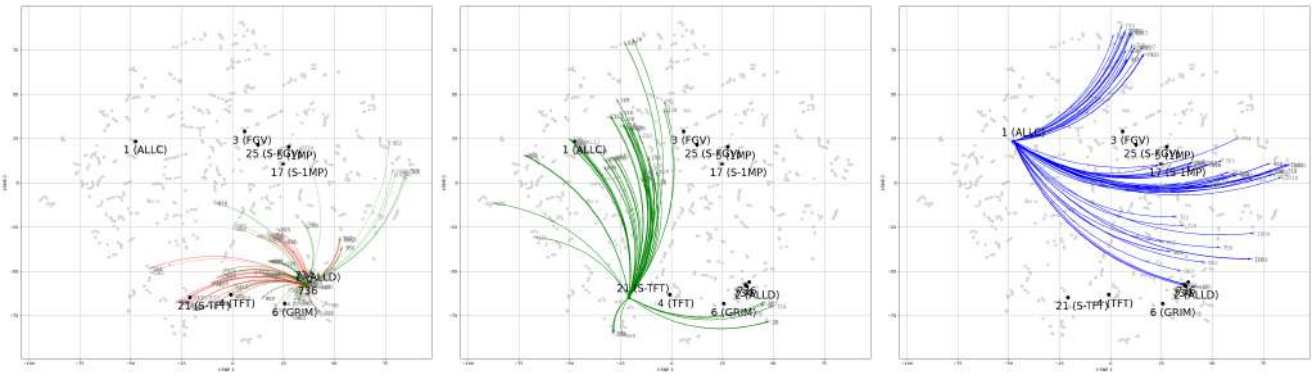
これに対し、#734 および#736 は S-TFT に対して相対的に頑健であり、淘汰されにくいことが図 8(a) および図 9(b) から読み取れる。S-TFT 自体はさまざまな戦略に淘汰されるものの、特に ALLC に淘汰されやすい性質がある。これは、S-TFT が初手で裏切る性質を持つのに対し、ALLC は常に協力を維持するため、S-TFT と ALLC の対戦において S-TFT が高い利得を得られにくく、ALLC が多数派を占める状況では淘汰圧を受けやすいためである。また、ALLC 同士の相互作用では高利得が安定的に得られるため、S-TFT が少数派である限りは淘汰されやすく、固定確率の観点からも不利な立場に置かれる。このような構造により、図 8 および 9 では S-TFT が ALLC へと淘汰されるアークが多く観察される。

ただし、ALLC もまた脆弱であり、ALLC に一度淘汰された S-TFT の周辺から、再び#734 や#736 の近傍へと進化が進むパターンが多く見られる。これは、ALLC に一時的に占有された集団に対して、より洗練された罰戦略（2 期相互処罰型）が進入し、再び淘汰を進めるという段階的な進化の構造を示唆している。このように、単純な裏切り戦略（ALLD）が S-TFT に淘汰され、さらに S-TFT が ALLC に淘汰され、そして最終的に ALLC がより精緻な処罰戦略に置き換えられてい



(a) ALLD および#736　(b) S-TFT　(c) ALLC

図 8: 見間違え環境下における 3 状態戦略の進化 ($g = l = 0.2$)



(a) ALLD および#736　(b) S-TFT　(c) ALLC

図 9: 見間違え環境下における 3 状態戦略の進化 ($g = l = 0.8$)

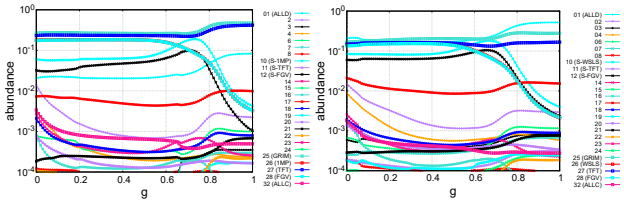
くという、淘汰の段階性を明らかにしている。これは、はじめにで述べたように、裏切りから始まり協調の維持を高度に制御する戦略が、進化動学の過程において最終的に安定して集団を支配するという本研究の主張を裏付けるものである。

6　ゲイン g に関する感度分析

本章では、ゲイン g の影響を吟味する。図 10-13 にロス l を 0.2 または 0.8 に固定して、ゲイン g を $[0.00, 1.00]$ の範囲で変化させたときの戦略比率を示す。図 10 および図 12 では g を 0.01 刻みで変化させ、図 11 および図 13 では g を 0.05 刻みで変化させた。横軸は g, 縦軸は戦略比率の対数を表す。また、図 10 および図 12 には 26 個の 2 状態以下の FSA 戦略を示しており、図 11 および図 13 には各 (g, l) の組で戦略比率が高い上位 5 戦略の戦略比率を示している。その他のパラメータを $N = 100, s = 10, \delta = 0.90$ とした。

6.1　行動の見間違えが起こらない場合

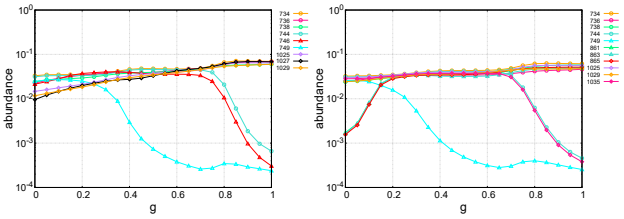
行動の見間違えがないとき、戦略空間を 2 状態以下の FSA 戦略に限定した場合、図 10 に示すようにロス l の大きさの影響はそれほど見られない。ゲイン g が小さいとき、最大多数戦略となる ALLD や GRIM 以外に TFT や 1MP, FGV がそれぞれ 10% 以上のシェアをもつ。ゲイン g が増加すると、ALLD や GRIM, TFT が生き残るようになる。ここで TFT が最大多数戦略にならないが、一定のシェアを持つのは TFT のどんな戦略と対戦しても大きく負けることはないという特徴を表している。



(a) $l = 0.2$ のとき　(b) $l = 0.8$ のとき

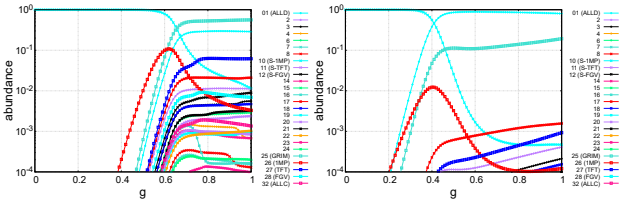
図 10: 戦略空間を 2 状態以下の FSA 戦略に限定した場合の行動の見間違えが起こらない場合のゲイン g に対する収束時の戦略比率

次に、戦略空間を 3 状態以下の FSA 戦略に拡張したときの



(a) $l = 0.2$ のとき　(b) $l = 0.8$ のとき

図 11: 戦略空間を 3 状態以下の FSA 戦略に限定した場合の行動の見間違えが起こらない場合のゲイン g に対する収束時の戦略比率



(a) $l = 0.2$ のとき　(b) $l = 0.8$ のとき

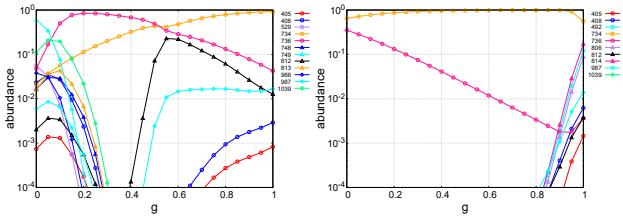
図 12: 戦略空間を 2 状態以下の FSA 戦略に限定した場合の行動を見間違えうる場合のゲイン g に対する収束時の戦略比率

結果を図 11 に示す。図 10 と同様に多くの戦略が広く薄く存在し、ゲイン g の大きさの影響はそれほど見られないが、一部の戦略はゲイン g の変化の影響が見られる。図 11a では、#744, #746, #749 の戦略比率はゲイン g が大きくなると低くなる。図 11b では、#744, #749, 1035 の戦略比率はゲイン g が大きくなると低くなり、#861 と#865 の戦略比率はゲイン g が大きくなると高くなる。

6.2　行動を見間違えうる場合

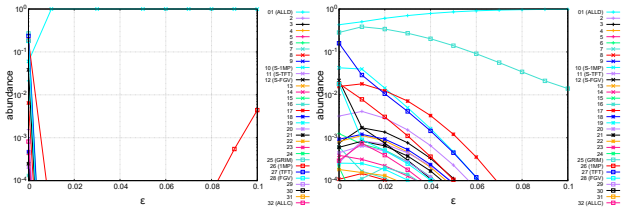
行動の見間違えがないとき、戦略空間を 2 状態以下の FSA 戦略に限定した場合、ロス l が小さいとき、図 12a に示すように、ゲイン g が十分小さい (約 0.65 以下) と S-1MP が、g が大きくなると GRIM が最大多数となる。一方、ロス l が大きくなると、図 12b に示すように、ゲイン g が十分小さい (約 0.40 以下) ときは S-1MP が変わらず最大多数となるが、g が大きくなると ALLD と GRIM がシェアを獲得するようになる。いずれもゲイン g が大きくなると S-1MP がそのシェアを急速に失う。

次に、戦略空間を 3 状態以下の FSA 戦略に拡張した場合、ロス l が小さいとき、図 13a に示すようにゲイン g が小さい (約 0.50 以下) と#736 が、g が大きくなると#734 が最大多数となる。一方、ロス l が大きくなると、図 13b に示すように #734 が常に最大多数となる。3 状態以下の FSA 戦略からなる戦略空間では、図 12 で生き残った戦略はほぼ生き残らなくなった。



(a) $l = 0.2$ のとき　(b) $l = 0.8$ のとき

図 13: 戦略空間を 3 状態以下の FSA 戦略に限定した場合の行動を見間違えうる場合のゲイン g に対する収束時の戦略比率



(a) $g = l = 0.2$ のとき　(b) $g = l = 0.8$ のとき

図 14: 戦略空間を 2 状態以下の FSA 戦略に限定した場合のプレイヤが行動を見間違える確率 ε に対する収束時の戦略比率

7　行動を見間違える確率 ε に関する感度分析

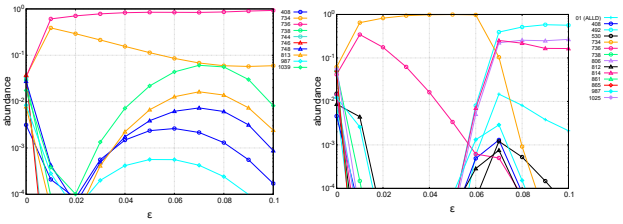
本章では、プレイヤが行動を見間違える確率 ε の影響を吟味する。図 14 および図 15 にゲイン g とロス l を 0.2 または 0.8 に固定して、ε を $[0.00, 0.10]$ の範囲で 0.01 刻みで変化させたときの戦略比率を示す。横軸は ε, 縦軸は戦略比率の対数を表す。また、図 14a および図 14b には 26 個の 2 状態以下の FSA 戦略の戦略比率を示しており、図 15a および図 15b には各 ε において戦略比率が高い上位 5 戦略の戦略比率を示している。なお、その他のパラメータを $N = 100, s = 10, \delta = 0.90$ とした。

7.1　2 状態

戦略空間を 2 状態以下の FSA 戦略に限定したとき、g と l が小さい ($g = l = 0.2$) と図 14a に示すように S-1MP が $\varepsilon = 0$ のときを除いて最大多数戦略となり、ほぼ 100% の戦略比率を獲得する。一方、g と l が大きくなると、図 14b に示すように ALLD が最大多数戦略となる。ε の増加に伴い、ALLD の戦略比率は高くなり、その他の戦略は低くなる。GRIM は ε が小さい範囲では ALLD に次いで高い戦略比率を獲得するが、ε が大きくなるにつれて ALLD 以外の戦略と同様に GRIM の戦略比率は低くなる。また、g と l の値に関わらず、$\varepsilon = 0$ のとき、つまり、見間違えが起こらないときは様々な戦略が広く薄く存在する。

7.2　3 状態

次に、戦略空間を 3 状態以下の FSA 戦略に拡張すると、g と l が小さいとき、図 15a に示すように $\varepsilon = 0$ のときは図 14 と同様にさまざまな戦略が広く薄く存在するが、ε が 0 より



(a) $g = l = 0.2$ のとき　(b) $g = l = 0.8$ のとき

図 15: 戦略空間を 3 状態以下の FSA 戦略に限定した場合のプレイヤが行動を見間違える確率 ε に対する収束時の戦略比率

大きくなると、主に#736 と#734 が生き残る。ε が大きくなると、#736 の戦略比率は高くなるが、#734 の戦略比率は緩やかに低くなる。また、#408, #748, #813, #987, #1039 の 5 戦略の戦略比率は ε が 0.02 を超えると ε の増加に伴って大きくなるが、ε が 0.07 を超えるとその戦略比率は低下する。一方で、g と l が大きいと ε が小さいときは#734 が最大多数戦略となり、ほぼ 100% に近い戦略比率を獲得するようになるが、ε が 0.06 を超えるとその戦略比率は急速に低下する。ε が大きいときは #492, #806, #814 の 3 戦略が生き残るようになる。これらの戦略はいずれも GRIM と似た構造を持つ戦略であり、シグナル b を観測するとその後は永遠に裏切り続ける構造を含んでいる。g と l が大きくなったことで裏切る誘因や裏切られることによる損失が大きくなったことに加え、ε が大きくなったことで協力を維持することが難しくなったため、GRIM と似た構造を持つ#492, #806, #814 が生き残るようになったと考えられる。

8　おわりに

本研究では、見間違えのあるくり返し囚人のジレンマにおける戦略の進化を、確率動学にもとづいて分析した。特に、有限状態機械によって定義された状態数 3 以下の戦略空間において、見間違えが戦略の進化に与える影響を明らかにした。見間違えのない環境では、ALLD や GRIM などの単純な非協力戦略が支配的になる一方、見間違えが導入されると、協調を維持するための洗練された罰戦略が進化の過程を通じて優勢になることが示された。

本研究の主な貢献は、固定確率にもとづく確率動学の分析により、裏切りから始まり段階的に進化していく戦略の経路を可視化した点にある。特に、ALLD が S-TFT に淘汰され、S-TFT が ALLC に淘汰され、最終的に#734 や#736 といった S-2MP 型戦略が集団を支配するという淘汰の段階性が確認された。また、t-SNE を用いて高次元の淘汰構造を視覚的に理解する手法は、進化の力学を定性的に捉える上で有効であることが示唆された。今後の課題としては、状態数のさらなる拡張や、異なるノイズモデル、報酬構造、戦略学習モデルへの応用が挙げられる。

参考文献

[1] Robert Axelrod. *Genetic Algorithms and Simulated Annealing*, chapter The Evolution of Strategies in the Iterated Prisoner's Dilemma, pages 32–41. Morgan Kaufman, Los Altos, CA, 1987.

[2] Seung Ki Baek, Hyeong-Chai Jeong, Christian Hilbe, and Martin Nowak. Comparing reactive and memory-one strategies of direct reciprocity. *Scientific Reports*, 6:25676, 2016.

[3] Christian Hilbe, Krishnendu Chatterjee, and Martin Nowak. Partners and rivals in direct reciprocity. *Nature Human Behavior*, 2:469–477, 2018.

[4] Michihiro Kandori. Repeated games. In Steven N. Durlauf and Lawrence E. Blume, editors, *Game theory*, pages 286–299. Palgrave Macmillan, 2010.

[5] G. Mailath and L. Samuelson. *Repeated Games and Reputation*. Oxford University Press, 2006.

[6] Philippe Mathieu and Jean-Paul Delahaye. New winning strategies for the iterated prisoner's dilemma. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '15, pages 1665–1666, Richland, SC, 2015. International Foundation for Autonomous Agents and Multiagent Systems.

[7] Martin Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, 2006.

[8] Martin Nowak and Karl Sigmund. A strategy of win-stay, lose-shift that outperforms tit for tat in prisoner's dilemma. *Nature*, 364:56–58, 1993.

[9] Karl Sigmund. *The Calculus of Selfishness*. Princeton University Press, 2010.

[10] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[11] Benjamin Zagorsky, Johannes Reiter, Krishnendu Chatterjee, and Martin Nowak. Forgiver triumphs in alternating prisoner's dilemma. *PLOS ONE*, pages 1–8, 2013.

[12] ヨンジュン ジョ, 岩崎 敦, 神取 道宏, 小原 一郎, and 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**, 53(11):2445–2456, 2012.

[13] 関口 格. **経済セミナー増刊: ゲーム理論プラス, 「協調達成のための正しいお仕置きの仕方」**. 日本評論社, 2007.

[14] 大槻 久. 有限集団における進化ゲーム理論の発展. **統計数理**, 60(2):2–5, 2012.

[15] 岡田 章. **ゲーム理論 新版**. 有斐閣, 2011.

[16] 松島 斉. **ゲーム理論の新展開**. 勁草書房, 2002. 第 4 章: 「繰り返しゲームの新展開: 私的モニタリングによる暗黙の協調」, pp.89-114.