CF-001

取り違えのある繰り返し囚人のジレンマにおける協力のダイナミクス

村井　伸一郎 *　　岩崎 敦 *
Shinnchiro Murai　　Atsushi Iwasaki

概要

本研究では，繰り返し囚人のジレンマにおいて，プレイヤが行動を取り違えるとき，無限集団上のダイナミクスのもとでどんな戦略が生き残るかを吟味した．従来よく使われる戦略表現として，一期記憶戦略がある．これは昨日の自分と相手の行動から今日の自分の行動を決める戦略表現である．しかしこれでは，昨日自分が意図した行動を考慮しないため，有名なトリガー戦略，一度でも裏切りを観測したら二度と協力しない，を正しく表現できない．裏切りを観測したあとに 2 人が同時に行動を取り違えると，再び協力するようになってしまう．そこで本論文では，昨日自分が意図した行動を考慮した戦略表現として有限状態機械を採用し，状態数 2 以下の非同相な戦略を列挙した空間上で突然変異付きレプリケータダイナミクスの帰結を吟味した．その結果，協力を維持する仕組みが利得構造や割引因子といった環境に応じてどのように変化していくかを明らかにした．

1　はじめに

繰り返しゲームは，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 [1, 2, 3] であり，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. [4, 5]．2 人がまったく行動を取り違えないならば，常に裏切り（ALLD）や一度でも裏切られたら許さない（Grim-trigger，GRIM）といった非協力的な戦略しか生き残らないことが知られている [6]．しかし，実際の人間はしばしば行動を取り違えることがある．例えば，協力しようとしたが失敗してしまったり，サボったつもりがうまくいってしまったりすることが起こると考えるのは自然である．こうした行動の取り違えは進化ゲーム理論における重要な仮定であると考えられている．実際，こうした間違いがないと，お互いに協力することが進化的安定性を満たさないことが知られている [7].

人と人がどうのように協力する（しない）かを分析するには，多くの研究で囚人のジレンマが用いられる．囚人のジレンマはお互いに裏切ることが支配戦略となるゲームであるが，実際の人々はこのような状況でもしばしば協力を維持することがある．直接互恵性 (direct reciprocity) はこれに対する重要な説明の 1 つであるが，具体的にどんな戦略のもとで協力を維持しているかは必ずしも自明ではない [8, 9, 10]．例えば，繰り返しゲームの理論におけるフォーク定理は，協力的な均衡の存在を証明することはできる [2]．しかし，GRIM 以外のどんな戦略で協力的均衡を構成するかは明らかでない．一方で，進化ゲームでは，状況に適応できる戦略が生き残ると考える．このような自然淘汰のもとでも裏切り (ALLD) への誘引が強いため，協力的な戦略は有名な TFT も含めて生き残りにくい [11]．このため，ALLD や GRIM 以外の戦略がいつどのように生き残るかを明らかにすることは人工知能，経済学，生物学といった複数の研究分野にまたがる重要な問いになっている．

本研究では，プレイヤたちが一定の確率で意図した行動と異なる行動を取ってしまう，行動の取り違え (implementation errors) [12] が発生するとき，突然変異付きレプリケータダイナミクスの帰結がどうなるかを吟味する．行動の取り違えについては広範な先行研究があるが，その多くは戦略空間をかなり限定する，もしくは戦略自体を進化させるような閉じていない戦略空間 [13] を想定している．本研究では，戦略空間を閉じた形で定義しつつ，従来より多くの戦略を表現できる有限状態機械戦略の上でのダイナミクスを吟味する．

先行研究では，反応戦略 (reactive strategies) もしくは 1 期記憶戦略 (memory-one strategies) が閉じた戦略空間としてしばしば用いられる [14]．反応戦略では，今日の自分の行動を，昨日の相手の行動だけから決める．この戦略空間は ALLD や ALLC（常に協力），そして TFT を含むが GRIM を含まない．次に 1 期記憶戦略では，今日の自分の行動を，昨日の自分と相手の行動から決める．この戦略空間でもっとも有名な戦略は “勝ち残り，負け逃げ” (Win-Stay，Lose-Shift，WSLS) である．これはプレイヤが将来の利得を重視し，協力のコストが十分小さいときに生き残る．しかし，この戦略空間でも行動を取り違えることを考慮すると，GRIM を含むことができない．実際，お互いに協力したときのみ協力し，それ以外が起こった後では必ず裏切るように設定すればよさそうに見える．このとき，お互いが裏切るとその後協力することはなさそうに見える．しかし，お互いが行動を取り違えて相互協力が発生すると再び相互協力に戻ることができてしまう．これは 1 期記憶戦略が昨日，自分が意図した行動と実際にとった行動を区別できていないためである．

そこで，本研究では戦略を有限状態機械 (Finite State Automaton，FSA) で表現する．FSA もよく使われる戦略表現であるが，自分が行動を取り違えたかどうかがわからないという

* 電気通信大学大学院情報理工学研究科

制限を課していた [15]．この結果，従来研究が吟味していたのは相手の行動を見間違える状況における 26 個の戦略であった [6]．本研究ではこの制限を緩めて，自分の意図した行動と実現した行動を区別することで，482 個の戦略からなる空間を扱う．

繰り返しゲームの戦略は，昨日までの履歴から今日の選択する行動への写像で定義する．ゲームを無限回繰り返すとき，その戦略空間は無限になるので，すべての均衡戦略を具体的に特定することは現実的ではない．そこで本論文では，プレイヤが取りうる戦略を状態数 2 以下の FSA に限定する．戦略を FSA に限定したときの期待利得は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突然変異付きレプリケータ方程式 [10] を解く．レプリケータダイナミクスとは，利得が高くなる戦略をとるプレイヤの人口は増加させ，低くなる戦略をとる人口はより良い戦略へ取って代わられてやがて絶滅するといった具合に自然淘汰の過程を表現する頻度依存淘汰モデルである [16, 17]．その結果，協力するコストが大きいときは ALLD が生き残り，協力するコストが小さくなるにつれて GRIM の変種が多く生き残る状態を経由して，WSLS の変種が複数生き残ることを明らかにした．また，自分が裏切る誘引が相手の裏切りによる損失よりも大きいとき，TFT が生き残りやすくなり，自分が裏切る誘引が小さい，かつ相手の裏切りによる損失が中程度の大きさであるとき，GRIM の変種と TFT 変種が常に協力する ALLC，もしくは WSLS の変種といった寛容な戦略と共存し，混合戦略均衡となることがわかった．また，GRIM の変種，WSLS の変種，および混合戦略均衡が生き残る領域は相互協力の実現頻度がほぼ等しく，行動を取り違えた後の振る舞いが協力状態を実現させやすくしていることを発見した．

2　モデル

2.1　行動の取り違えのモデル

本章では行動の取り違えのある無限回繰り返しゲームをモデル化する．ここでプレイヤ $i \in \{1,2\}$ はステージゲームを無限期間 $t = 0,1,2,\cdots$ に渡って繰り返す．割引因子は $\delta \in (0,\ 1)$ とする．各期においてプレイヤ i は有限集合 $A_i = \{C,\ D\}$ から行動 a_i を選択し，その行動の組を $\boldsymbol{a} = (a_1, a_2) \in A^2$ とする．このとき，意図した行動を $\bar{a}$，実現した行動を $\hat{a}$ とする．プレイヤ 1 の利得 $g_1(\hat{a})$ は利得表によって定められた値に従う．また，意図した行動の組 $(\bar{a}_1, \bar{a}_2)$ に対して，実現した行動の組 $(\hat{a}_1, \hat{a}_2)$ が生起する同時確率を $o((\hat{a}_1, \hat{a}_2)|(\bar{a}_1, \bar{a}_2))$ とする．

本論文では利得表として表 1 に示す囚人のジレンマを用いる．表中の C は協力行為を，D は裏切り行為を表す．囚人のジレンマの利得構造は $g > 0$, $l > 0$ であり，このとき D は厳密な支配戦略となる．また、囚人のジレンマでは $|g - l| < 1$ が要求される．もしこの条件が成り立たないとすると，繰り返し囚人のジレンマにおいて協力と裏切りを交互に出すほうが，純粋な協力よりも利得が高くなってしまい，純粋な協力が維持できなくなる．

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

	$\hat{a}_2 = C$	$\hat{a}_2 = D$
$\hat{a}_1 = C$	$1,\ 1$	$-l,\ 1+g$
$\hat{a}_1 = D$	$1+g,\ -l$	$0,\ 0$

表 2:　同時確率分布 $o((\hat{a}_1, \hat{a}_2)|(\bar{a}_1, \bar{a}_2))$

	$\hat{a}_2 = \bar{a}_2$	$\hat{a}_2 \neq \bar{a}_2$
$\hat{a}_1 = \bar{a}_1$	p	q
$\hat{a}_1 \neq \bar{a}_1$	q	$1-p-2q$

次に，両プレイヤが実現した行動が意図した行動と一致した ($\hat{a}_1 = \bar{a}_1$かつ$\hat{a}_2 = \bar{a}_2$) とき，$o((\hat{a}_1, \hat{a}_2)|(\bar{a}_1, \bar{a}_2)) = p$ とする．また，片方のプレイヤが実現した行動が意図した行動と一致しなかった ($\hat{a}_1 \neq \bar{a}_1$かつ$\hat{a}_2 = \bar{a}_2$，もしくは $\hat{a}_1 = \bar{a}_1$かつ $\hat{a}_2 \neq \bar{a}_2$) とき，$o((\hat{a}_1, \hat{a}_2)|(\bar{a}_1, \bar{a}_2)) = q$ とする．p が最も高くなるように設定し，この同時確率分布を表 2 に示す．

2.2　FSA 戦略

繰り返しゲームの戦略は，昨日までの履歴から今日の選択する行動への写像で定義する．本研究では行動を取り違えた後の振る舞いを網羅し，状態数 2 以下の非同相な 482 個の FSA を戦略空間とする．FSA の状態は R (reward, 報酬) と P (punishment, 処罰) の 2 つに区別され, プレイヤ i は状態 R で行動 $a_i = C$ を選び, 状態 P で行動 $a_i = D$ を選ぶ．それぞれの状態でプレイヤは自分と相手がとった行動で次にどの状態に遷移するかが決まる．簡単のため，実現した行動の組が (C, C) のときは CC，(C, D) のときは CD，(D, C) のときは DC，(D, D) のときは DD と表す．例えば，状態 R からは 4 つの行動の組に対して状態遷移が決まる．図 2 の各 FSA において，状態 R における CC や CD は自分が行動を取り違えなかったときの，DC や DD は自分が行動を取り違えたときの遷移を表している．

行動の取り違えにおいて FSA を列挙すると 482 個にものぼる．このため，行動の取り違えがないときの振る舞いを基準に戦略を図 1 のように分類する．状態数 1 の戦略には ALLD と ALLC が存在し，ALLD (図 1a) は状態 P のみを持ち毎期必ず裏切る戦略，ALLC (図 1b) は状態 R のみを持ち毎期必ず協力する戦略である．一方で状態 R と P を持つ状態数 2 の著名な戦略としてはトリガー戦略 (Grim-trigger，GRIM, 図 1c) と “しっぺ返し”(Tit-For-Tat，TFT，図 1d) がある．GRIM は，状態 R からスタートして最初は協力し，一度でも相手が裏切ると，それ以降永遠に裏切り続ける戦略である．TFT は，状態 R からスタートし，相手の協力を観測した次の期には協力を，裏切りを観測した次の期には裏切りを行う戦略である．他にも重要な戦略として，“勝ち残り，負け逃げ” (Win-Stay, Lose-Shift，WSLS，図 1e) が存在する．WSLS は，状態 R からスタートし，プレイヤは最初に協力し，相手が裏切るとプレイヤも裏切るが，互いに 1 期裏切った後，そのプレイヤは協力に戻る．最後に，裏切られたら一度だけ相手を処罰し協力へ戻る戦略である “Forgiver”（FGV，図 1f）が存在する．FGV は状態 R からスタートし，相手の裏切りを観測した次の期のみ裏切るが，その後は何を観測しても協力に戻る戦略である．

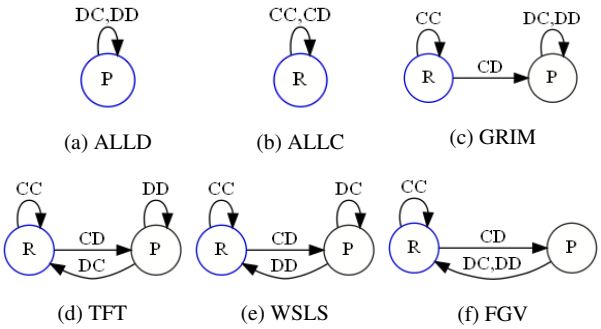


図 1: 繰り返しゲームにおける重要な戦略

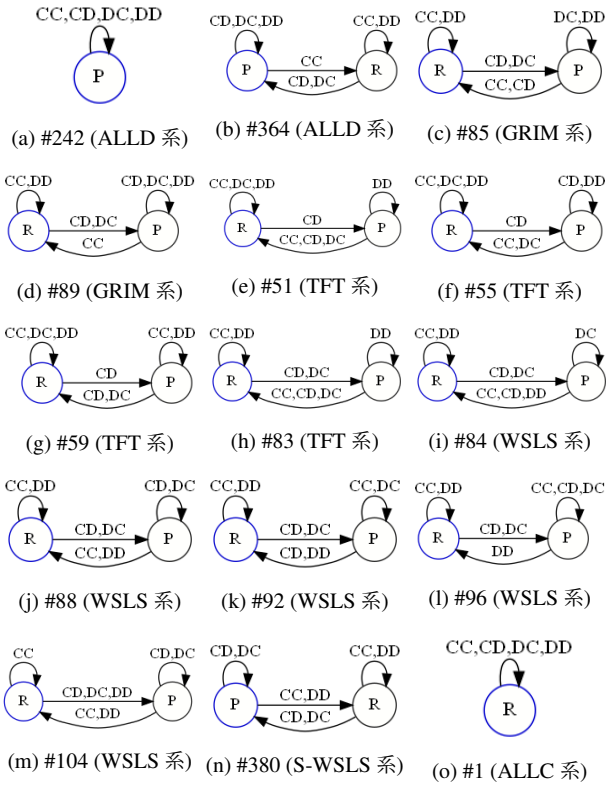


図 2: 行動の取り違えにおける戦略の例

以上の 6 つの戦略の系にしたがって 482 戦略を分類すると，ALLD 系もしくは ALLC 系に含まれるのはそれぞれ 49 戦略ずつであり，他の 4 つの戦略の系にはそれぞれ 16 戦略ずつが含まれる．

図 2 に本研究で定義した 482 戦略のうち 15 個の戦略を示す．詳しくは 4 章以降で述べるが，比較的広い範囲のパラメータ下，一定の割合で生き残る戦略となっている．

まず，ALLD 系の戦略から 2 戦略を挙げる．#242 (図 2a) は自分と相手の行動によらず常に裏切ろうとする戦略である．状態 R に遷移することが無いため，1 状態の戦略として表現できる．#364 (図 2b) は後述する#89 と初期状態のみが異なる戦略である．

次に，GRIM 系の戦略から#85 (図 2c) と#89 (図 2d) について示す．#85 と#89 は共に状態 R で自分が行動を取り違えて裏切ったとき，相手が協力すると次の期は状態 P に遷移して裏切ろうとし，相手も裏切ると次の期は状態 R にとどまって協力しようとする．違いとして，#85 は状態 P で自分が行動を取り違えて協力すると次の期は必ず状態 R に戻って協力しようとする戦略である．一方，#89 は状態 P では自分が行動を取り違えて協力したとき，相手が協力すると次の期は状態 R に戻って協力しようとし，相手が裏切ると状態 P にとどまって裏切ろうとする戦略である．

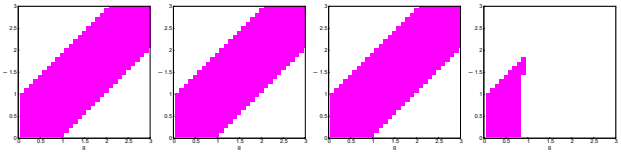
TFT 系の戦略から 4 つの戦略を説明する．まず，#51 (図 2e)，#55 (図 2f) および#59 (図 2g) は状態 R で行動を取り違えたとき，次の期も必ず協力しようとし，状態 P で行動を取り違えたときの振る舞いが異なる．状態 P で自分が行動を取り違えると，#51 は次の期は必ず協力しようとする戦略である．#55 は相手が協力すると次の期は協力しようとし，相手が裏切ると裏切ろうとする戦略である．#59 は相手が協力すると次の期も裏切ろうとし，相手が裏切ると協力しようとする戦略である．次に，#83 (図 2h) は状態 R で自分が行動を取り違えたとき，相手が協力すると次の期は裏切ろうとし，相手が裏切ると次の期は協力しようとする．状態 P では自分が行動を取り違えると次の期は必ず協力しようとする戦略である．

WSLS 系の戦略から 5 戦略について示す．まず，#84 (図 2i) と#88 (図 2j)，#92 (図 2k) と#96 (図 2l) の 4 戦略は状態 R で自分が行動を取り違えたとき，相手が協力すると次の期は裏切ろうとし，相手が裏切ると協力しようとする戦略である．この 4 戦略は状態 P で行動を取り違えたときの振る舞いが異なる．まず，#84 は状態 P で自分が行動を取り違えると，次の期に必ず協力しようとする戦略である．#88 は相手が協力すると次の期は協力しようとし，相手が裏切ると裏切ろうとする戦略である．#92 は相手が協力すると次の期も裏切ろうとし，相手が裏切ると協力しようとする戦略である．#96 は相手の行動によらず次の期も裏切ろうとする戦略である．#104 (図 2m) は状態 R で自分が行動を取り違えると次の期は必ず裏切ろうとする．状態 P で行動を取り違えたとき，相手が協力すると次の期は協力しようとし，相手が裏切ると裏切ろうとする戦略である．また，重要な戦略として#380 (図 2n) があり，これは行動を取り違えないときの振る舞いが状態 P から始まる WSLS (Suspicious WSLS, S-WSLS) となる戦略である．この戦略は#88 と初期状態だけが異なり，振る舞いは同じ戦略である．

最後に，ALLC 系の戦略である#1 (図 2o) は自分と相手の行動によらず常に協力しようとする．状態 P に遷移することが無いため，1 状態の戦略として表現できる．

2.3 ナッシュ均衡とサブゲーム完全均衡

ここではナッシュ均衡とサブゲーム完全均衡について概説する．ナッシュ均衡とは，ゲーム開始時点でみたとき，お互いに最適な戦略を取り合っている状態のことである [2]．また，



(a) #242 (b) #364 (c) #89 (d) #96

図 3: ナッシュ均衡 ($p = 0.95, q = 0.01$)

相手がある FSA にしたがって振る舞うとき，自分の割引利得和を最大化する FSA を最適反応 FSA と呼ぶ．ある FSA の組がお互いに最適反応 FSA となっているとき，その FSA の組はナッシュ均衡になっていることと等価である [18]．図 3 にナッシュ均衡を構成する戦略の例と利得パラメータ g と l についてプロットした図を示す．図の横軸は利得パラメータ g, 縦軸は l を表しており，$p = 0.95, l = 0.01$ で固定した．図 3a- 3c からわかるように，#242，#364，#89 の 3 戦略はどの利得パラメータの組においてもナッシュ均衡となる．また，図 3d より，#96 は $g \leq 0.80$ では常に，$g = 0.90$ では $l \geq 1.50$ でナッシュ均衡となる．

一方，サブゲーム完全均衡とは，どんなことが起こった後でも常に戦略で指定された行動を取ることがお互い最適になっている状態である [2]．本研究では行動を取り違えた後の振る舞いを網羅した戦略空間を扱っているため，サブゲーム完全均衡であるかを判定することにより，行動を取り違えた後の振る舞いが最適な戦略とそうでない戦略を区別できる．図 2 で示した戦略の中では，#242, #364, #89 が全ての利得パラメータの組において，#96 が $g \leq 0.90$ のときにサブゲーム完全均衡となる．

表 3: $p = 1.00, q = 0.00$ における利得表 ($g = 0.10, l = 0.10$)

Strategy	#242	#364	#89	#85	#83	#51	#55	#96	#1
#242	$\underline{0.000}^{*,\dagger}$	$\underline{0.000}^{*}$	0.110	0.110	0.110	0.110	0.110	0.579	$\underline{1.100}$
#364	$\underline{0.000}^{*}$	$\underline{0.000}^{*,\dagger}$	0.110	0.110	0.110	0.110	0.110	0.579	$\underline{1.100}$
#89	-0.010	-0.010	$\underline{1.000}^{*,\dagger}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\overline{1.000}$
#85	-0.010	-0.010	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\overline{1.000}$
#83	-0.010	-0.010	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	1.000
#51	-0.010	-0.010	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\overline{1.000}$
#55	-0.010	-0.010	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\overline{1.000}$
#96	-0.053	-0.053	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*}$	$\underline{1.000}^{*,\dagger}$	$\overline{1.000}$
#1	$\overline{-0.100}$	$\overline{-0.100}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	$\underline{1.000}$	1.000

表 3 に行動を取り違えることがないパラメータ $p = 1.00, q = 0.00, g = l = 0.10$ における利得表を示す．簡単のため，列の戦略に対する行の戦略の利得のみを示している．表中の下線（上線）は列（行）の戦略に対する最適反応を，$*$ はその戦略の組が 482 戦略間でのナッシュ均衡になっていることを表す．また，サブゲーム完全均衡となる戦略の組は † で示した．$p = 1.00, q = 0.00$ のとき，ALLD 系の戦略をとると，自身の利得は負になることはない．また，相手が ALLD 系の戦略であるとき，他の戦略を選ぶと利得は必ず負になる．また，ALLD 系の戦略が ALLC 系の戦略と対戦するとき，他のどの戦略が対戦するときよりも高い利得を得ることができる．そのため，ALLC 系の戦略がナッシュ均衡になることはない．GRIM 系，TFT 系および WSLS 系の戦略同士の対戦では常にナッシュ均衡を構成する．

2.4 突然変異付きレプリケータダイナミクス

本研究のように，数ある戦略の中から有効な戦略を発見する方法の 1 つとして，レプリケータダイナミクスがある．ゲームを行うプレイヤの集団を考え，プレイヤはいくつかの戦略の中からランダムに戦略を選択し，他のプレイヤとゲームを行い利得を得る．その後，戦略の集団に対する利得と集団全体の平均利得との差に応じて戦略の人口比を増減させる [12]．本論文では，上述のレプリケータダイナミクスに突然変異の概念を加える．突然変異付きレプリケータダイナミクスでは，適応度による人口の変化に加えて，すべての戦略が適応度に関係なく一定の確率で異なる戦略をとるとする．したがって，ある戦略が突然変異する確率を u とおき，突然変異付きレプリケータ方程式を

$$\dot{x_i} = x_i \left[f_i(\vec{x}) - \phi(\vec{x}) \right] + u \left(\frac{1}{n} - x_i \right), \ i = 1, ..., n$$

と定義する [15]．$\phi(\cdot)$ を全ての戦略の利得の平均 $\sum_j x_j f_j(\vec{x})$，$f_j(\cdot)$ を $\sum_j x_j a_{jm}$ とする．ただし，a_{jm} は戦略 j をとるプレイヤが戦略 m を取るプレイヤと無限回プレイしたときの割引利得和である．

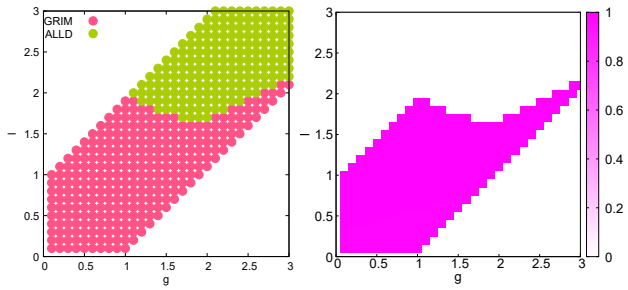
2.5 実験設定

数値実験では，割引利得 ($\delta = 0.90$) を固定した上で，g, l を $[0.1, 3.0]$ の範囲で 0.1 刻みで変化させた．戦略として状態数 2 以下の非同相な 482 個の FSA を用いる．また，初期時点において，各戦略の人口は一様に分布，つまり，各戦略の存在割合は全て等しいものとする．さらに，突然変異を起こす確率 u を 0.01 とした．また，50000 期で計算を終了し，帰結が収束しなかった場合は 40001 期目から 50000 期 の平均を評価した．

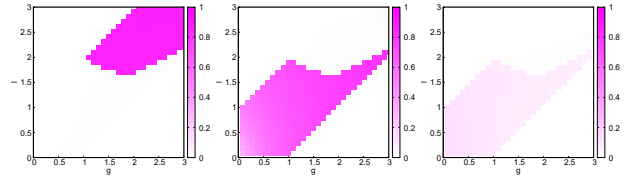
3 行動の取り違えが無いとき

図 4 に行動の取り違えが無いときにおけるレプリケータダイナミクスの帰結を示す．ここでは，同時確率分布のパラメータを $p = 1.00, q = 0.00$ とした．それぞれの図の縦軸は相手の裏切りによる損失 l，横軸は自分の裏切りによる利得の増分 g に対応し，0.1 刻みで $[0.1, 3.0]$ をプロットした．図 4a に最大多数戦略を，図 4b に協力率を示す．最大多数戦略とは，収束時に最も多くの人口を獲得した戦略を意味し，協力率とは，収束時の戦略分布に対する CC の実現頻度のことである．図 4b より，ALLD 系の戦略が最大多数となるときの協力率はほぼ 0 であるが，GRIM 系の戦略が最大多数となるときはほぼ 1.00 である．これは，ALLD 系のような非協力的な戦略は必ず裏切り合うのに対して，GRIM 系，TFT 系，FGV 系，WSLS 系，および ALLC 系の戦略は恒久的な協力関係を築くことができるためである．

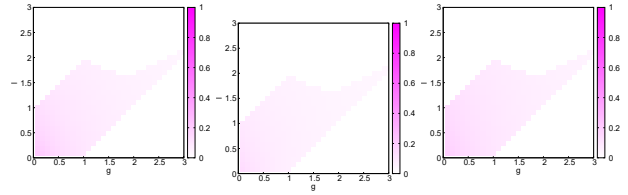
また，残りの図 4c- 4h は収束時に生き残った戦略の割合を



(a) 最大多数戦略　(b) 協力率



(c) #242 および#364 と等価な 49 戦略 (ALLD 系)　(d) #89 および#85 と等価な 16 戦略 (GRIM 系)　(e) #51, #55 および #83 と等価な 16 戦略 (TFT 系)



(f) #96 と等価な 16 戦略 (WSLS 系)　(g) FGV 系の 16 戦略　(h) #1 と等価な 16 戦略 (ALLC 系)

図 4: 行動の取り違えが無いときのダイナミクス

示している．図 4c は，#242 (図 2a) および#364 (図 2b) と等価な 49 戦略，つまり ALLD 系全ての戦略が収束時に生き残る割合の合計値を示している．同様に，図 4d は GRIM 系全ての戦略，図 4e は TFT 系全ての戦略，図 4f は WSLS 系全ての戦略，図 4h は ALLC 系全ての戦略の割合の合計である．また，図 4 には示していないが，FGV 系の戦略も生き残る．図 4a から見てもわかるように，行動の取り違えが無いときは，主に生き残るのは不寛容な戦略であることがわかった．

ゲイン g とロス l が大きいときは ALLD 系の戦略，それ以外の領域では GRIM 系の戦略が最大多数となる．例として，$g = 2.00, l = 2.00$ では ALLD 系の全 49 戦略が生き残る割合の合計値は 0.867 となる．一方で，GRIM 系の戦略が最大多数となるとき全 16 戦略合計で 0.300〜0.800 ほどの割合で生き残り，GRIM 系の戦略は TFT 系，WSLS 系，FGV 系および ALLC 系と共存する．共存する割合は g と l の値に依存し，g と l が大きくなるにつれて GRIM 系の戦略が占める割合は増加する．例として，$g = l = 0.01$ では，収束時の割合が (GRIM 系, TFT 系, WSLS 系, FGV 系, ALLC 系) = (0.309, 0.136, 0.213, 0.136, 0.190) となり，$g = l = 0.10$ では，(GRIM 系, TFT 系, WSLS 系, FGV 系, ALLC 系) = (0.663, 0.091, 0.084, 0.060, 0.092) となる．

表 4: $p = 0.95, q = 0.01$ における主要 9 戦略間の利得表 ($g = 0.10, l = 0.10$)

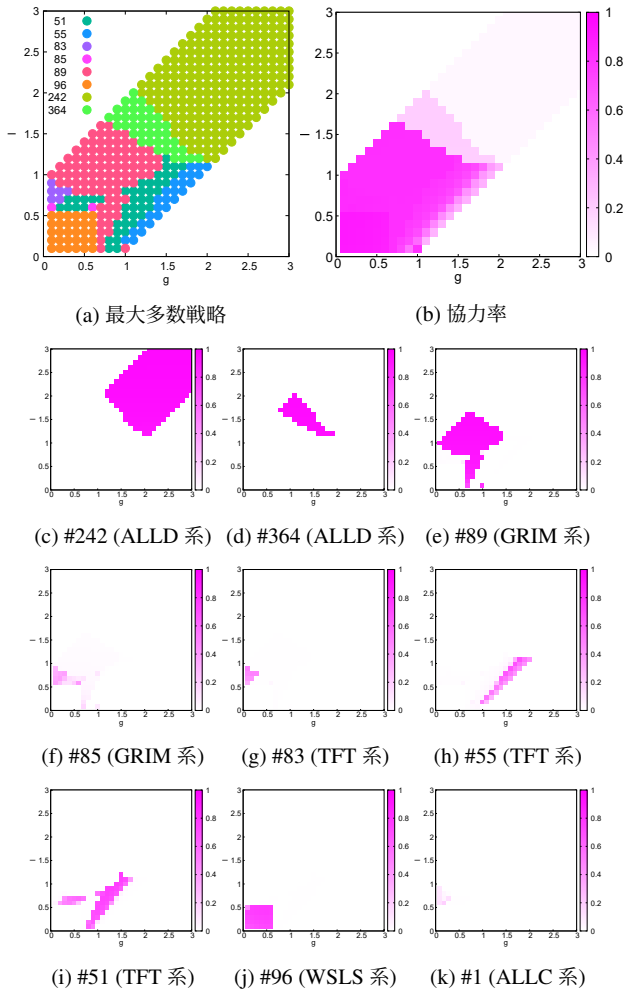
Strategy	#242	#364	#89	#85	#83	#51	#55	#96	#1
#242	$\overline{0.040}^{*,\dagger}$	0.067	0.167	0.175	0.183	0.187	0.179	0.566	$\underline{1.052}$
#364	0.038	$\overline{0.211}^{*,\dagger}$	0.302	0.308	0.314	0.318	0.312	0.574	1.045
#89	0.028	0.191	$\overline{0.846}^{*}$	0.847	0.848	0.855	0.854	0.892	0.973
#85	0.028	0.189	0.846	0.847	$\overline{0.853}$	0.862	0.861	0.891	0.971
#83	0.027	0.188	0.845	0.852	0.858	0.883	0.879	0.869	$\overline{0.962}$
#51	0.027	0.187	0.844	0.852	0.878	0.906	0.901	0.866	$\overline{0.961}$
#55	0.027	0.188	0.844	0.852	0.874	0.902	0.895	0.867	$\overline{0.961}$
#96	-0.008	0.043	0.820	0.823	0.867	0.873	0.872	$\overline{0.943}^{*,\dagger}$	0.973
#1	$\overline{-0.052}$	0.020	0.817	0.842	$\underline{0.942}$	$\underline{0.951}$	$\underline{0.951}$	0.817	0.960

4　行動の取り違えがあるとき

本節ではプレイヤが意図した行動と異なる行動に取り違えうる時のダイナミクスの帰結を示す．この実験は 482 個もの戦略を含むため，全ての戦略を取り上げることはできない．そこで，ダイナミクスの収束時における割合が 0.4 を超える利得パラメータが少なくとも 1 つ存在した戦略として 8 個の戦略に ALLC 系の戦略#1 を加えて主要 9 戦略を定義した．表 4 にこの 9 戦略同士からなる利得表を示し，図 2 に対応する戦略を示している．表 4 は簡単のため，列側の戦略に対する行側の戦略の利得のみを示している．表中の下線（上線）は列（行）の戦略に対する最適反応を表す．さらに $*$ はその戦略の組が 482 戦略間でのナッシュ均衡になっていることを，$\dagger$ はそのサブゲーム完全均衡となっていることをそれぞれ表している．

図 5 に行動の取り違えがあるときにおけるレプリケータダイナミクスの帰結を示す．同時確率分布のパラメータは $p = 0.95, q = 0.01$ とした．図 5a に最大多数戦略を，図 5b に協力率を示す．また，残りの図 5c- 5k は収束時における主要戦略の割合を示している．図 5a が示すように，どんな戦略が生き残るかは利得構造に依存し，傾向としては g と l が小さくなるにつれて，最大多数戦略が協力的になることがわかった．また，図 5a と図 5b を比較すると最大多数戦略は異なるが，協力率がほぼ同じ領域が存在する．これは生き残るための振る舞いはかなり異なっても，平均的には同じような頻度で協力を達成していることを示す．

図 5a に示すように裏切ることによるゲイン g と裏切られることによるロス l が十分大きいとき，常に裏切ろうとする戦略 #242 (図 2a) が他の戦略を支配する．例えば，$g = l = 2.00$ のとき，#242 は 0.952 の割合で生き残る．このとき，2 人のプレイヤが同時に行動を取り違えて協力する確率 0.03 がそのまま協力率となる．ここから協力するコストやリスクが小さくなるとき，徐々に協力的な戦略が生き残るようになる．例えば，$g = l = 1.50$ になると，#364 (図 2b) が 0.953 の割合で生き残るようになる．これは ALLD 系の戦略であるが，初期状態である状態 P で行動を取り違えて協力したとき，相手も協力していたら，状態 R に遷移するようになっている．さらにその後はお互いに協力もしくはお互いに裏切る限りは，状態 R に



(a) 最大多数戦略　(b) 協力率

(c) #242 (ALLD 系)　(d) #364 (ALLD 系)　(e) #89 (GRIM 系)

(f) #85 (GRIM 系)　(g) #83 (TFT 系)　(h) #55 (TFT 系)

(i) #51 (TFT 系)　(j) #96 (WSLS 系)　(k) #1 (ALLC 系)

図 5: 行動の取り違えがあるときのダイナミクス

留まる．その結果，協力率が 0.195 に増加する．さらに g と l が小さくなると，今度は#89 (図 2d) が最大多数を占めるようになり，$g = l = 1.00$ では 0.935 の割合で生き残るようになる．#89 は GRIM 系の戦略であるが，#364 とは初期状態のみが異なる戦略である．#89 同士が対戦するとき，共に状態 P にいるときに同時に取り違えることで協力状態に戻るタイミングを明確にし，その後の相互協力を実現しやすくしている．このため，協力率は 0.819 とかなり高い水準に到達する．

次に，WSLS 系の戦略である#96 (図 2l) が最大多数となるまで g と l が十分小さくするときを考える．例えば，$g = l = 0.10$ では#96 が生き残る割合は 0.659 にとどまった．このとき，他の WSLS 系の戦略である#92 (図 2k) および#88 (図 2j) がぞれぞれ 0.124 および 0.058 の割合で生き残る．図 6a にこのときのダイナミクスの時間変化を示しており，これら 3 つの戦略が急速に 8 割以上のシェアを獲得していることがわかる．この WSLS 系の戦略が生き残る領域での協力率は 0.899 と非常に高くなる．WSLS 系同士の対戦では，どちらか一方のプレイヤが行動を取り違えて協力状態が途切れた後も，互いに裏切り合う相互処罰を経て，協力状態に簡単に戻ることができる．互いに罰を与えることで相互協力に戻るのは，一見直感に反するが，相互処罰がうまく協力に戻るタイミングを明確にしている．

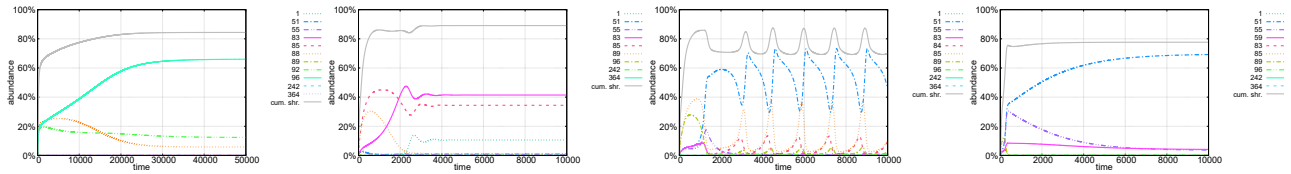
次に裏切られることによるロス l が大きすぎなく，かつ小さすぎない場合を考える．図 5a が示すように，このときはある特定の戦略の系が支配的になることはなくなる．裏切ることによるゲイン g によってそのダイナミクスは変化する．まず，g が十分小さいとき，GRIM 系の#85 (図 2c)，TFT 系の#83 (図 2h) および ALLC 系の#1 (図 2o) の 3 戦略が共存する．実際 $g = 0.10, l = 0.70$ のとき，そのダイナミクスは図 6b のように変化し，(#85, #83, #1)=(0.344, 0.414, 0.106) に収束する．また，#85 と#83 はいずれも状態 P における取り違えから協力を回復する戦略であり，その協力率は 0.833 になる．

次に g を 0.7 にまで増加させると GRIM 系の#85，TFT 系の#51 (図 2e) および WSLS 系の#84 (図 2i) の 3 戦略がサイクルを形成するようになる．最後の 10000 期の結果を平均すると，その割合は (#85, #51, #84)=(0.106, 0.553, 0.045) になり，その協力率は 0.817 になる．図 6c に $g = 0.70, l = 0.70$ におけるダイナミクスの変化を示しているが，早い段階でこれら 3 戦略がサイクルを形成している．

さらに g を大きくしていくと，ゲイン g とロス l の差分の絶対値が徐々に 1 に近づいていく．このとき，TFT 系である#51 もしくは#55 が最大多数となる．例えば，図 6d に $g = 1.30, l = 0.70$ でのダイナミクスの変化を表している．ここでは，#51 が最大多数となるとともに#59 (図 2g) と#55 (図 2f) が小さい割合で生き残るようになり，その収束時の割合は (#51, #59, #55)=(0.696, 0.040, 0.035) となる．これらの 3 戦略は全て TFT 系であり，行動を取り違えるとき，相互に協力と裏切りを繰り返すようになる．ここで g と l の差分の絶対値が比較的小さいときは，相互協力による利得と協力と裏切りを繰り返すことによる利得が近くなる．その結果，TFT 系が生き残るようになる．このため，その協力率は 0.736 と GRIM 系の戦略（例えば#89）が最大多数となるときよりも低くなる．また，$g = 1.60, l = 0.70$ では#55 が最大多数となるが，生き残った割合は 0.407 と多くない．代わりに 0.010 程度の割合で生き残る戦略を多数観察した．

5　感度分析

本章では，取り違えの同時確率分布，利得，割引因子といったゲームのパラメータに関する感度を分析する．まず，2 人のプレイヤが行動を取り違えない確率を表す p を変化させる．ここでいずれか 1 人のプレイヤが行動を取り違える確率 $q = 0.10$ とし，利得パラメータ g および l を 0.10 に，割引因子 δ を 0.90 に固定する．このとき，図 7 に p を $[0.50, 1.00]$ の範囲で 0.02 刻みで変化させたときの最大多数戦略の期待利得を示す．p が大きくなる，つまり取り違えが起こりにくくなるにつれて，最大多数戦略は#380 (図 2n) , #89 (図 2d) , #88 (図 2j) , #96 (図 2l) , #104 (図 2m) と変わっていき，実現する期



(a) $g = 0.10, l = 0.10$　(b) $g = 0.10, l = 0.70$　(c) $g = 0.70, l = 0.70$　(d) $g = 1.30, l = 0.70$

図 6: ダイナミクスの時間変化

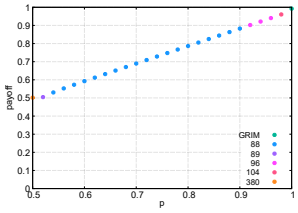


図 7: 取り違えない確率 p

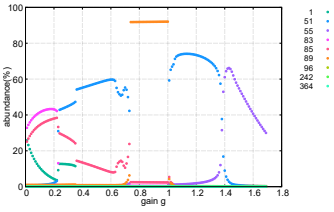


図 8: ゲイン g

待利得が増加する．$p = 1.00$ にしたときは例外的に $q = 0.00$ とし，取り違えが起きないようにした．この結果，GRIM 系に属する 49 個の戦略が一様に分布しながら最大多数になる．

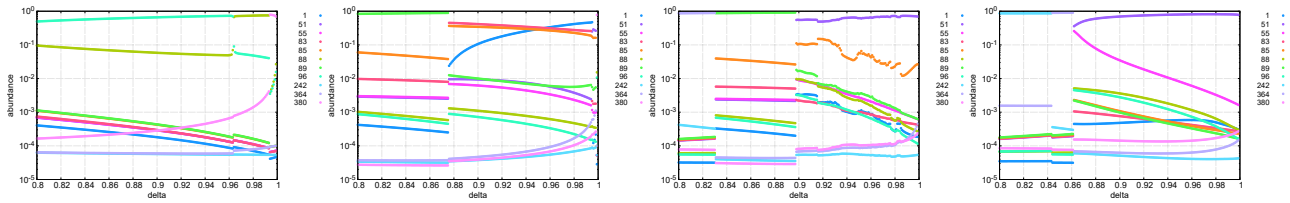
$p = 0.95$ としていた図 5a は，利得パラメータ g および l を 0.10 に固定すると WSLS 系の#96 が最大多数戦略となっていた．この利得パラメータで p を動かすと， #96 は $p \in [0.91, 0.99]$ で最大多数戦略になり，同じく WSLS 系の #88 が $p \in [0.54, 0.90]$ で最大多数戦略となる．#96 より#88 の方がより頑健な戦略であることがわかる．詳細は省略するが，これは行動を取り違えた後の振る舞いから生じる利得が #96 より#88 の方が高いためである．

次に，図 8 に $l = 0.7$ に固定し，裏切ることで得る利得の増分 (ゲイン) g に対して主要な戦略の収束時の割合がどのように変化するかを示す．ここで横軸は利得パラメータ g を 0.01 刻みで $[0.01, 1.69]$ の範囲で変化させた．縦軸は各戦略の収束時の割合とした．他のパラメータは $p = 0.95, q = 0.01, \delta = 0.90$ とした．g が十分小さい ($g \in [0.00, 0.22]$) 時，どれが最大多数戦略になるか変わるが，図 6b で見たように，GRIM 系の#85 (図 2c) , TFT 系の#83 (図 2h) , および ALLC 系の#1 (図 2o) の 3 戦略が共存する．#83 が最大多数であるが，#85 もほぼ同程度の割合で生き残り，g が大きくなるにつれて，#83 と#85 は割合を増加させる一方で，#1 の割合は単調に減少する．次に $g \in [0.23, 0.35]$ では，TFT 系の戦略が#83 から #51 (図 2e) に変化し，#51 が#85 と#1 と共存する．やがて，g が 0.36 を越えると#1 が絶滅する．さらに g が 0.74 を越えてくると GRIM 系の#89 が 0.900 以上の割合で他の戦略を支配する．最後に，g が 1.00 以上になると，TFT 系の戦略である #51 と#55 (図 2f) が共存して他の戦略を支配するようになる．TFT 系の戦略では一方が行動を取り違えると，協力と裏切りを交互に繰り返す，すなわち CD, $DC, CD, \ldots$ となる．プレイヤが互いに協力と裏切りを繰り返すとき，各期の利得の平均は $\frac{1+g-l}{2}$ となる．もしここで $g - l = 1$ であれば，$\frac{1+g-l}{2} = 1$ となる．これは永遠に相互協力を続ける時の平均利得に一致する．よって，ゲインとロスの差が 1 に近いときは，TFT 系の戦略でも相互協力並みの利得を維持でき，他の戦略に対して大負けしないという性質と合わせて，生き残ると考えられる．

最後に，割引因子がダイナミクスの帰結に与える影響を吟味する．割引因子はプレイヤが将来の利得をどれだけ割り引いて考えるかを表すパラメータである．割引因子が高ければ高いほど，プレイヤは将来の利得を重要視する，つまり相手の裏切りに対して我慢強く振る舞い，将来的な協調を実現させようとする．図 9 に割引因子 δ を $[0.800, 0.999]$ の範囲を 0.001 刻みで動かしたときの主要な戦略の割合を示す．ここでは $p = 0.95, q = 0.01,$ に固定し，主要 9 戦略に#88 と#380 を追加した 11 戦略の割合を示した．図 9a に $g = l = 0.10$ としたときの結果を示す．ここでは，割引因子 δ が 0.96 以下のとき，WSLS 系の#96 が最大多数を占める．δ がさらに大きくなり 0.96 を越えると同じ WSLS 系の#88 が最大多数になり，0.994 を越えると状態 P から始まる WSLS である S-WSLS に準じて振る舞う#380 に切り替わる．

図 9b に $g = 0.10, l = 0.70$ としたときの結果を示す．ここでは割引因子が 0.875 以下のとき，GRIM 系の#89 が単独で他の戦略を支配する．続いて割引因子を増加させていくと，GRIM 系の#83 および#85 に ALLC 系の#1 を加えた 3 つの戦略が共存するようになる．さらに割引因子が 1 に近づくと，他の戦略に負けにくくなる TFT 系の#51 から S-WSLS 系の#380 が最大多数戦略となる．このように割引因子が大きくなり，プレイヤが将来利得を重要視するようになると，状態 P から状態 R に戻りやすい戦略が生き残るようになる．実際，#89 は CC が実現するときだけ状態 R に戻るが，#83 や#85 は CD や DC が実現しても状態 R に戻るようになっている．さらに割引因子が 1 に近づくと，協力に戻るためにコストをかけても将来利得で補償されるので，TFT 系や S-WSLS 系といった非自明な戦略で協力を維持することができるようになる．

図 9c に $g = l = 0.70$ としたときの結果を示す．ここでは割引因子が 0.83 以下のとき，GRIM 系の#364 が，0.89 以下のときは GRIM 系の#89 が単独で他の戦略を支配する．さらに 0.89 を越えると，TFT 系の#51 が最大多数を占める．このとき，わずかに GRIM 系の#85 が生き残っているもののその割合は 0.1 を越えることはほとんどない．ここでも割引因子が



(a) $g = 0.10, l = 0.10$　(b) $g = 0.10, l = 0.70$　(c) $g = 0.70, l = 0.70$　(d) $g = 1.30, l = 0.70$

図 9: 割引因子 δ の影響

大きくなると TFT 系の戦略が有利になる現象が観察できた．しかし，GRIM 系の戦略が最大多数になるときと比べるとその割合はあまり安定しない．

最後に，図 9d に $g = 1.30, l = 0.70$ としたときの結果を示す．ここでは割引因子が 0.84 以下のとき，ALLD 系の#242 (図 2a) が，次いで ALLD 系の#364 が最大多数を占め，他の戦略を支配する．割引因子が 0.86 を越えると TFT 系の#51 と #55 が共存するようになり，徐々に#51 が単独で他の戦略を支配するようになる．ここでは $g - l$ が 0.6 と 1 に近づいており，割引因子が小さいと近視眼的な ALLD 系が支配的になるものの，プレイヤが将来利得を十分考えるようになると，TFT 系の戦略に切り替わる．$g = 1.30, l = 0.70$ では裏切りによる利得の増分と損失が大きく，短期的な利得の損失を抑えるため，δ が小さいときには#242 もしくは#364 といった不寛容な戦略が単独で支配する．

6　おわりに

本研究では，取り違えのある繰り返し囚人のジレンマを突然変異付きレプリケータダイナミクスで分析した．もしプレイヤが行動を取り違えることがないと，ALLD 系の非協力的な戦略もしくは GRIM 系の相手の裏切りを二度と許さない不寛容な戦略のいずれかが最大多数を占める．一方で，プレイヤが行動を取り違えるとき，協力のコストやリスクが小さくなるにつれて ALLD 系から協力的な戦略が生き残るようになる．例えば，同じ GRIM 系でも取り違えから相互協力を観測したら協力に戻る戦略が生き残ったり，相互処罰を通じて協力に戻る WSLS 系の戦略が生き残る．これらの戦略における協力率は同じ水準を保っており，協力のコストやリスクに応じて取り違えた後の振る舞いが変化することを明らかにした．

参考文献

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

[2] 神取道宏. 人はなぜ協調するのか –くり返しゲーム理論入門–. 三菱経済研究所, 2015.

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

[4] 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.

[5] 岡田章. ゲーム理論 新版. 有斐閣, 2011.

[6] 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. 情報処理学会論文誌, Vol. 63, No. 4, pp. 1138–1148, apr 2022.

[7] Drew Fudenberg and Eric Maskin. Evolution and cooperation in noisy repeated games. *The American Economic Review*, Vol. 80, No. 2, pp. 274–279, 5 1990.

[8] David G. Rand and Martin A. Nowak. Human cooperation. *Trends in Cognitive Sciences*, Vol. 17, No. 8, pp. 413 – 425, 2013.

[9] Julián García Matthijs van Veelen, David G Rand, and Martin A. Nowak. Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences*, Vol. 109, No. 25, pp. 9929–9934, 2012.

[10] Lorens A. Imhof, Drew Fudenberg, and Martin A. Nowak. Evolutionary cycles of cooperation and defection. *in Proceedings of the National Academy of Sciences*, Vol. 102, No. 31, pp. 10797–10800, 2005.

[11] Martin A. Nowak. Five rules for the evolution of cooperation. *Science*, pp. 1560–1563, 2006.

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

[13] Huanren Zhang. Errors can increase cooperation in finite populations. *Games and Economic Behavior*, Vol. 107, No. C, pp. 203–219, 2018.

[14] Chatterjee K. Hilbe, C. and M.A. Nowak. Partners and rivals in direct reciprocity. *Nat Hum Behav 2*, p. 469–477, 2018.

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

[16] Peter D. Taylor and Leo B. Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, pp. 145–156, 1978.

[17] Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[18] 神取道宏. ミクロ経済学の力. 日本評論社, 2014.