CF-009

# 取り違えのある繰り返し囚人のジレンマにおける単独裏切-相互処罰戦略

村井 伸一郎 *　　岩崎 敦 *
Shinnchiro Murai　　Atsushi Iwasaki

## 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] 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 [3]．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永遠に裏切り続ける無期限罰則のトリガー戦略 (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 [4]．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 [5].

人と人がどのように協力する（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 [6]．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しっぺ返し (Tit-For-Tat, TFT) 戦略も含めて生き残りにくい [7]．このため，ALLD や GRIM 以外の戦略の生存過程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) [8] が起こりうるとき，突然変異付きレプリケータダイナミクスの帰結を吟味する．行動の取り違えに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閉じていない戦略空間 [9] 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，従来より多くの戦略を表現できる有限状態オートマトン戦略の上でのダイナミクスを吟味する．

従来，閉じた戦略空間としては 1 期記憶戦略 (memory-one strategies) がしばしば用いられる [10]．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，協力するコストが十分小さいときに生き残ることが知られている．

本研究では，1 期記憶戦略を拡張した有限状態オートマトン戦略の空間に焦点を当てる．具体的には，戦略を非同相な有限状態オートマトン (Finite State Automaton，FSA) として列挙する．これにより 1 期記憶戦略では表現できない，自分が意図した行動と実現した行動の違いを考慮した戦略を表現できる．進化ゲームの文脈でも有限状態オートマトンを用いた戦略表現が採用されているが，1 期記憶戦略しか列挙できていなかったり，自分が行動を取り違えたかどうか観察できないという制限を課していたりしていた [11]．したがって本研究は，行動を取り違える環境下で，プレイヤが自分が意図した行動と実現した行動の違いを考慮して振る舞う状況を吟味する．

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，プレイヤが取りうる戦略を状態数 2 以下の有限状態オートマトンに限定する．戦略を有限状態オートマトン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 [6] 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 [12].

その結果，“単独裏切-相互処罰” (Unilateral Defection, Mutual Defection, UDMD) という新しい戦略を発見した．この戦略は，従来よいとされていた WSLS とよく似た戦略であるが，プレイヤが行動を取り違えにくく，将来の利得をあまり重視しないとき，WSLS を淘汰することがわかった．UDMD も WSLS も協力のコストが小さいときに生き残る戦略だが，UDMD の方がより広範なパラメータの下で WSLS や他の戦略を淘汰し，生き残る．また，実際に UDMD の方が WSLS よりも均衡になりやすいことを理論的に明らかにした．

* 電気通信大学大学院情報理工学研究科

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 $o(\hat{\boldsymbol{a}} \mid \boldsymbol{a})$

| | $\hat{a}_2 = a_2$ | $\hat{a}_2 \neq a_2$ |
|---|---|---|
| $\hat{a}_1 = a_1$ | $p$ | $q$ |
| $\hat{a}_1 \neq a_1$ | $q$ | $s$ |

## 2 モデル

### 2.1 行動の取り違えのモデル

本節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 $t = 0, 1, 2, \ldots$ 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，意図した行動の組 $\boldsymbol{a}$ に対して，実現した行動の組を $\hat{\boldsymbol{a}} = (\hat{a}_1, \hat{a}_2) \in A^2$ とする．各期におけるプレイヤ $i$ の利得を成分ゲームの利得関数 $g_i(\hat{\boldsymbol{a}})$ で与える．また，意図した行動の組 $\boldsymbol{a}$ に対して，実現した行動の組 $\hat{\boldsymbol{a}}$ が生起する同時確率を $o(\hat{\boldsymbol{a}} \mid \boldsymbol{a})$ とする．成分ゲームの利得関数として，表 1 に示す囚人のジレンマを用いる．ここで，$C$ は協力行為を，$D$ は裏切り行為を表す．利得パラメータとして，相手が協力しているとき，自分が裏切ることで得る利得の増分（ゲイン）$g > 0$ および自分が協力しているとき，相手に裏切られることで得る利得の減分（ロス）$l > 0$ を用いる．行動 $D$ を成分ゲームの支配戦略とするため，$|g - l| < 1$ を仮定する．

次に，囚人のジレンマに合わせて $o(\hat{\boldsymbol{a}} \mid \boldsymbol{a})$ を定義する．まずお互いの意図した行動と実現した行動が一致する確率を $p$ とする．いずれか一方のプレイヤだけ意図と実現した行動が一致する確率を $q$ とし，お互いの意図と実現した行動が一致しない確率を $s = 1 - p - 2q$ とする．このような同時確率分布を表 2 にまとめる．

繰り返しゲームの戦略は，全てのプレイヤの $t$ 期までの行動の履歴を記録し，$h^t = (\boldsymbol{a}^0, \boldsymbol{a}^1, \ldots, \boldsymbol{a}^t) \in H^t \equiv (A \times A)^{t+1}$ とする．各プレイヤの初期行動 $\boldsymbol{a}$ を決定するためのダミーとして $h^0$ を導入する．ここで $h^0$ は単一集合 $\{h^0\}$ とする．次にプレイヤ $i$ の純粋戦略 $s_i$ を，あらゆる履歴をある行動に対応させる関数として定義する．厳密には，ありうる履歴の集合 $H = \bigcup_{t \geq 0} H^t$ に関して，$s_i : H \to A$ とする．無限期間の繰り返しゲームを考えるとき，履歴の集合の大きさも無限大になるため，その戦略を簡略に表記する必要がある．そこで，有限状態オートマン (finite state automaton) がよく用いられる．あるオートマトン $m$ を状態の集合 $\Theta$, 初期状態 $\hat{\theta} \in \Theta$, 各状態で選択される行動 $f : \Theta \to A$, 決定的状態遷移 $T : \Theta \times A \times A \to \Theta$ に対して，$\langle \Theta, \hat{\theta}, f, T \rangle$ と定義する．ここで，決定的状態遷移 $T(\theta^t, \hat{\boldsymbol{a}}^t)$ は現在の状態 $\theta^t$ および実現した行動の組 $\hat{\boldsymbol{a}}^t$ に対して，次の期の状態 $\theta^{t+1}$ を返す関数とする．

戦略 $s_i$ をオートマトン戦略 $m_i$ で表現する．戦略の組 $\boldsymbol{m} = (m_1, m_2)$ にしたがって行動するプレイヤ $i$ の利得を割引利得和で定義する．2 人のプレイヤが任意の状態にいるときの，割引利得和を

$$V^{\boldsymbol{m}}_{\theta_i, \theta_{-i}} = \sum_{\hat{\boldsymbol{a}} \in A^2} o(\hat{\boldsymbol{a}} | (f(\theta_i), f(\theta_{-i}))) g_i(\hat{\boldsymbol{a}}) + \delta \sum_{\hat{\boldsymbol{a}} \in A^2} o(\hat{\boldsymbol{a}} | (f(\theta_i), f(\theta_{-i}))) V^{\boldsymbol{m}}_{T(\theta_i, \hat{\boldsymbol{a}}), T(\theta_{-i}, \hat{\boldsymbol{a}})} \quad (1)$$

とする．行動を取り違えうる繰り返しゲームの均衡として，まずナッシュ均衡を定義する．

**定義 1.** オートマトン戦略の組 $\boldsymbol{m}$ がナッシュ均衡であるとは，任意のプレイヤ $i$ の所与の初期状態における割引利得和が，戦略 $m_i$ からの任意の戦略 $m_i'$ への逸脱によって改善しないことである．すなわち

$$V^{\boldsymbol{m}}_{\hat{\theta}_i, \hat{\theta}_{-i}} \geq V^{(m_i', m_{-i})}_{\hat{\theta}_i', \hat{\theta}_{-i}}, \text{ for all } m_i'.$$

ナッシュ均衡において，どんな戦略に逸脱するかを通常は制限しないが，この定義では各プレイヤは任意のオートマトン戦略にしか逸脱しないものとする．オートマトン戦略で均衡を記述する限り，戦略空間を限定した均衡概念しか定義できないという訳ではない．そこで，戦略空間を限定せずにオートマトン戦略のペアが均衡になる概念を導入する．そのため，任意のオートマトン戦略 $m_i$ に対して，各状態で定められた行動をとり，実現した行動にもとづいて状態遷移した後，$m_i$ にしたがって行動する戦略 $\bar{m}_i$ を考える．このような一度だけ逸脱して後は元の戦略に従うような戦略を一回逸脱戦略と呼び，利得を改善するような一回逸脱戦略が存在しないとき，その戦略の組を，戦略空間を限定しない均衡として保証できる．これを一回逸脱原理と呼ぶ [12].

**定義 2.** オートマトン戦略の組 $\boldsymbol{m}$ が完全公的均衡 *(perfect public equilibrium)* であるとは，任意のプレイヤ $i$ および任意の状態の組 $(\theta_i, \theta_{-i})$ における割引利得和が，戦略 $m_i$ からの任意の一回逸脱戦略 $\bar{m}_i$ への逸脱によって改善しないことである．すなわち

$$V^{\boldsymbol{m}}_{\theta_i, \theta_{-i}} \geq V^{(\bar{m}_i, m_{-i})}_{\bar{\theta}_i, \theta_{-i}}, \text{ for all } \bar{m}_i.$$

完全公的均衡は，行動を取り違える状況におけるサブゲーム完全均衡に相当する均衡概念である．行動を取り違えうるとき，2 人のプレイヤは実現した行動を正確かつ共通して観測するため，自分と相手が実際に取った行動を公的なシグナルとして解釈できる．

### 2.2 突然変異付きレプリケータダイナミクス

従来の均衡概念とは別に，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，進化ゲームの文脈でしばしば

用いられる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 [8]．本論文では，これに突然変異の概念を加えた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，適応度と関係のない一定の確率で各戦略が異なる戦略をとるとする．この突然変異確率を $u$ とおき，レプリケータ方程式を

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

と定義する [11]．$\phi(\cdot)$ を全ての戦略の利得の平均 $\sum_j x_j f_j(\vec{x})$，$f_j(\cdot)$ を $\sum_m x_j a_{jm}$ 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り，式 1 のベルマン方程式を解くことで得られる．

### 2.3 実験設定

数値実験では，割引因子 ($\delta = 0.90$) 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．戦略空間として，状態数 2 以下の非同相な 482 個のオートマトン戦略を用いる．また，初期時点において，各戦略の人口はランダムに生成し，全戦略の比率の和が 1 になるように正規化した．さらに，突然変異を起こす確率 $u$ を 0.01 とした．また，100000 期で計算を終了し，帰結が収束しなかった場合は 90001 期目から 100000 期の平均を評価した．

## 3 主なオートマトン戦略

本節では，繰り返しゲームにおいてよく知られている既存のオートマトン戦略について概説する．まず，図 1 に 6 つの 1 期記憶戦略 (memory-one strategies) を示す．1 期記憶戦略とは，昨日の自分と相手の行動から今日の行動を決める戦略のクラスであり，理論生物学分野でよく用いられている．いっけんオートマトン戦略と等価に見えるが，今日の自分がどの行動を意図するかを考慮していない点で異なる．言い換えると，1 期記憶戦略はある行動を観測した後の状態遷移が今日の状態に影響しないようになっている．

図 1c に示す無期限罰則のトリガー戦略 (Grim trigger, GRIM) は，最初は協力するが，一度でも裏切りを観測したら，それ以降（行動を取り違えない限りは）永遠に裏切る戦略である．しかし，本モデルでは行動を取り違えうるため，お互いが状態 $P$ にいるとき，お互いが行動を取り違えると協力に戻ることになる．つまり，行動の組 $CC$ が実現したとき，状態 $R$ と状態 $P$ のいずれにいても状態 $R$ に遷移して，協力するようになっている．（繰り返しゲームの理論が想定する）GRIM を厳密に記述するには，$CC$ が実現したとき，状態 $R$ もしくは状態 $P$ のいずれかにいるときはともにその状態に留まらなければならない．そのような戦略は 1 期記憶戦略のクラスには属さずオートマトン戦略の空間を考えなければならない．

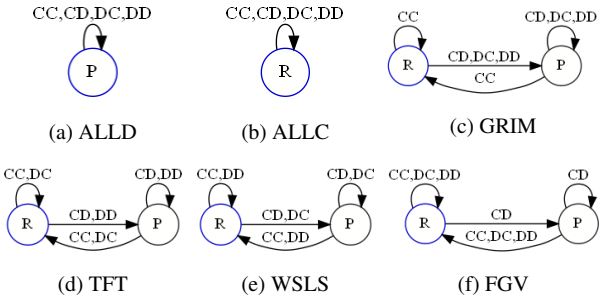
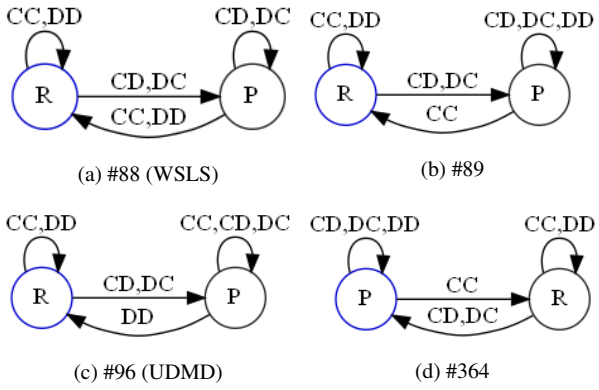



図 1: 繰り返しゲームにおける重要な戦略

図 1a および 1b は状態数が 1 個しかないため，自分がどちらの状態にいるかを考慮する必要はない．図 1d は非専門家の間でも有名なしっぺ返し (Tit-For-Tat, TFT) 戦略であり，図 1e は専門家の間では，よく知られている「勝ち残り，負け逃げ」(Win-Stay, Lose-Shift, WSLS) 戦略である．これについては 5 節で詳しく述べるが，行動の取り違えが起きうるとき，生き残ることがある 1 期記憶戦略は ALLD と WSLS のみである．図 1f は Forgiver と呼ばれる戦略で，自分が協力したにも関わらず裏切られたときのみ，裏切りを選択するが，それ以外では必ず協力する寛容な戦略である．これは行動を取り違えうるが，自分が行動を取り違えたかどうか認識できていない戦略空間において生き残りやすいことが知られている [11]．

本論文では，状態数 2 以下の非同相な 482 個のオートマトン戦略からなる戦略空間における突然変異付きレプリケータダイナミクスの帰結を分析する．図 2 に 482 個の戦略の中で生き残りやすかった 4 つの戦略を示す．図 2b および 2d に示す #89 と#364 は初期状態が異なるだけの同じ戦略である．#89 は最初状態 $R$ で協力し，$CC$ もしくは $DD$ が実現したら状態 $R$ に留まり，$CD$ もしくは $DC$ で状態 $P$ に遷移する．状態 $P$ に遷移したら，$CC$ が実現するまで状態 $P$ に留まる．これは図 1c とよく似た戦略であるが，状態 $R$ で $DD$ を観測しても状態 $R$ に留まる点で異なる．図 2a に示す#88 は図 1e と等価な戦略である．最初状態 $R$ で協力し，$CC$ もしくは $DD$ が実現したら状態 $R$ に留まり，$CD$ もしくは $DC$ で状態 $P$ に遷移する点は#89 と共通している．しかし，状態 $P$ で $CC$ もしくは $DD$ が実現することで状態 $R$ に遷移し，協力を回復する．この点で#88 は#89 より相手を許しやすく協力を回復させやすい戦略である．図 2c が本論文で新たに発見した “単独裏切-相互処罰 (Unilateral-Defection, Mutual-Defection, UDMD)” である．これは 4 節で詳しく述べる．

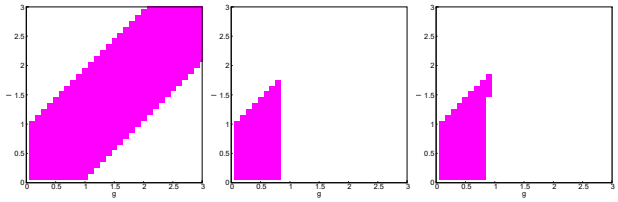
行動を取り違えることがなければ，482 個のオートマトン戦略のうち多くの戦略がナッシュ均衡を構成する．しかし，行動を取り違えうるとき，均衡になる戦略は限られる．具体的には，ALLD (#242), #89, #364, WSLS, UDMD の 5 つの戦略である．図 3 にナッシュ均衡を構成する戦略の例と利得パラメータ $g$ と $l$ についてプロットした図を示す．図の横軸は利得パラメータ $g$, 縦軸は $l$ を表しており，他のパラメータは




(a) #88 (WSLS) (b) #89

(c) #96 (UDMD) (d) #364

図 2: 行動の取り違えにおける戦略の例



(a) #89, #242, #364 (b) WSLS (#88) (c) UDMD (#96)

図 3: ナッシュ均衡 ($\delta = 0.9, p = 0.95, q = 0.01$)

$\delta = 0.9, p = 0.95$ $q = 0.01$ とした．図 3b および 3c に示すように，$l$ の値にほぼ依らず，WSLS は $g \leq 0.8$ で，UDMD は $g \leq 0.9$ でナッシュ均衡を構成する．

## 4 勝ち残り-負け逃げと単独裏切り-相互処罰

本節では，２つの重要な戦略，勝ち残り-負け逃げ (WSLS) と単独裏切-相互処罰 (UDMD) を分析する．WSLS は 1 期記憶戦略の中で，相互協力を基本としつつ，裏切った相手を報復でき，間違いを修正できる唯一の戦略である．図 1e に示すように，最初に状態 $R$ で協力し，$CC$ もしくは $DD$ が実現したら状態 $R$ に留まり，$CD$ もしくは $DC$ で状態 $P$ に遷移する．状態 $P$ では裏切り，$CD$ もしくは $DC$ が実現している限りは状態 $P$ に留まり，$CC$ もしくは $DD$ が実現したら状態 $R$ に遷移し，協力を回復する．お互いに裏切ったとしても，協力を選択するのはいっけん非合理的に見えるかもしれない．お互いが WSLS にしたがってプレイするとき，自分が状態 $R$ にいれば，その積状態は $RR$ もしくは $RP$ のいずれかである．$DD$ が実現したら，少なくとも自分が行動を取り違えたことがわかるので，積状態が $RR$ であれば，相手が状態 $R$ に留まることがわかる一方で，積状態が $RP$ であれば，相手が状態 $R$ に戻ることがわかる．このため，いずれの積状態であっても，$DD$ を観測した後，状態 $R$ に留まる方が相互協力を途切れさせることなく将来利得を維持できる．自分が状態 $P$ にいるときも同じことが言えるので，お互いに裏切ったとしても，次に協力を選択することが合理的になる．さらに，相手が意図的に裏切ったとしても，少なくとも 1 回は裏切ることで相手を処罰する．このため，WSLS は裏切りに対する罰を与えながらも協力を回復させやすい戦略になっている．その結果，1 期記憶戦略の空間では，相互協力が十分な利得を与えるとき，様々な設定で生き残ることが知られている．

本論文で初めて定義する UDMD は状態 $R$ からの遷移は WSLS と共通するが，状態 $P$ において $CC$ が実現した後は，状態 $R$ に遷移せず状態 $P$ にとどまる点が異なる．つまり，$CC$ が実現した後の振る舞いが状態 $R$ と状態 $P$ どちらの状態にいるかで異なる意味で，1 期記憶戦略に属さない戦略となっている．2 人のプレイヤが UDMD にしたがってプレイするとき，いずれかのプレイヤが裏切ると，積状態 $RR$ から $PP$ に遷移する．このとき，行動の取り違えが起こらなければ，つまり確率 $p$ で，$DD$ が実現し，すぐに $RR$ に復帰する．WSLS の場合，$CC$ が実現しても $RR$ に復帰できるので，$PP$ に遷移した後，確率 $1-2q$ で $RR$ に復帰できる．このため，WSLS の方が 2 人のプレイヤが同時に行動を取り違える確率 $s$ の分，$RR$ に復帰しやすい戦略となっている．逆に言うと UDMD の方がいったんお互いを処罰し始めると，WSLS より少しだけ処罰が強い戦略になっている．このことは，次に示すこれらの２つの戦略が均衡を構成する条件からも確認することができる．

**定理 1.** 割引因子を $\delta \in (0,1)$，利得パラメータを $g > 0$ および $l > 0$，行動を取り違える確率 $p > 0$ および $q > 0$ とする．ただし，$|g - l| < 1, p \gg q, p \gg s$ とする．このとき *WSLS* が完全公的均衡を構成するのは，

$$\delta \geq \frac{l(q-s) + g(p-q)}{(p+s-2q)(p-s)} \tag{2}$$

が成立するときである．

**定理 2.** 定理 1 と同じゲームパラメータを考える．*UDMD* が完全公的均衡を構成するのは，

$$\delta \geq \frac{l(q-s) + g(p-q)}{(p+s-2q)(p-s) + s\{l(q-s) + g(p-q)\}} \tag{3}$$

が成立するときである．

定理 1 および 2 は WSLS および UDMD がそれぞれ均衡を構成する割引因子の下限を示している．例えば，$p = 1$ および $q = 0$ のとき，Eq. 2 および Eq. 3 はともに $\delta \geq g$ となり，2 つの戦略の均衡条件は等しくなる．もし $q = s$ ならば，$p \gg q$ および $p \gg s$ より，Eq. 2 は $\delta \geq \frac{g}{p}$ に，Eq. 3 は $\delta \geq \frac{g}{p(1+sg)}$ となる．仮定より，Eq. 3 の右辺の分母の $s$ の係数 $l(q-s) + g(p-q)$ は正なので，Eq. 3 の分母は Eq. 2 の分母よりも大きく，均衡を構成する割引因子の下限は UDMD の方が小さい．このため，UDMD が完全公的均衡になるときは WSLS も均衡になるが，その逆は必ずしも成り立たない．一般にトリガー戦略のように処罰が強い戦略の方が均衡を構成しやすいので，この UDMD は WSLS より少しだけ均衡を構

表 3: 4 戦略の利得表 ($p = 0.95, 0.01, g = l = 0.1$)

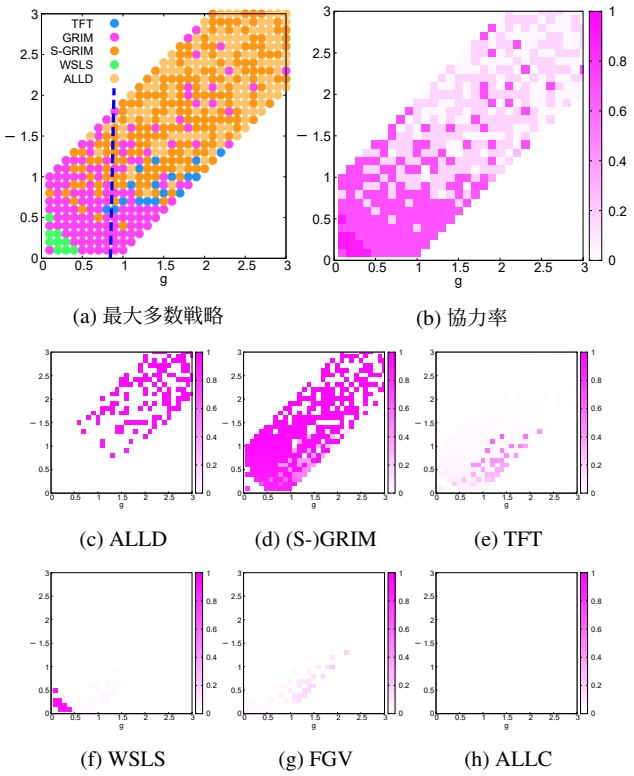
| Strategy | ALLD | UDMD | WSLS | ALLC |
|---|---|---|---|---|
| ALLD | 0.04 | 0.57 | 0.57 | 1.05 |
| UDMD | -0.01 | 0.94 | 0.94 | 0.97 |
| WSLS | -0.01 | 0.94 | 0.94 | 0.97 |
| ALLC | -0.05 | 0.82 | 0.83 | 0.96 |

成しやすいという結果は，UDMD が WSLS より少しだけ処罰が強い戦略という事実と整合している．

ここまで 1 期記憶戦略の中で優れているとされる WSLS と 1 期記憶戦略に属さない新しい戦略である UDMD の振る舞いの特徴を述べ，その完全公的均衡となる条件を導出し，議論した．WSLS も UDMD もともに利得パラメータ $g$ および $l$ が十分小さい，つまり協力のコストが十分小さければ，均衡を構成することを明らかにした．また，行動を取り違えないときは，2 つの戦略を均衡という観点からは区別できない一方で，行動を取り違える確率が大きくなると，ともに均衡を構成する割引因子の下限が大きくなり，将来利得を重視する我慢強いプレイヤでないと均衡を構成しなくなる．したがって，WSLS と UDMD の 2 つの戦略を均衡で区別するには，行動を取り違えうる必要があり，その帰結は取り違えの確率と割引因子によって変化する．

表 3 に ALLD, UDMD, WSLS, ALLC の 4 つの戦略がそれぞれ対戦した時の利得表を示す．ここで，利得と取り違え確率のパラメータは $p = 0.95, 0.01, g = l = 0.1$ とした．このとき，ALLD，UDMD, WSLS の 3 戦略は均衡を構成する．100 回プレイするうち，5 回ほど行動を取り違えるとき，お互いが無条件に協力すると，その平均利得（正規化した割引利得和）は 0.96 となる．これに対して，UDMD や WSLS は 0.94 の平均利得を均衡で達成しており，これらの戦略は，間違いが起きうる状況で，処罰のコストを抑えつつ相互協力を実現しているよい戦略と言える．しかし，先の均衡条件と同様，均衡利得の意味でも UDMD と WSLS を区別することはやはり難しい．

そこで，これら 2 つの戦略の間のダイナミクスを考えてみよう．UDMD と WSLS はともに均衡を構成するので，ある不動点となる戦略の分布が存在して，そこからわずかでもずれるとずれた方の戦略が支配的になる（不安定な均衡点）．例えば，突然変異がない場合，WSLS の比率が 0.55 より多くなると WSLS が支配的になり，少なくなると UDMD が支配的になる．突然変異率を増加させると，その不動点の位置はほとんど変わらないが，ダイナミクスの帰結における両戦略の比率は変化する．WSLS と UDMD の 2 つの戦略しかないとき，これら 2 つの戦略の間で突然変異することになるため，突然変異率が十分大きいと 2 つの戦略の比率は一様ランダム，つまりお互いに等しく比率を分け合うことになる．したがって，突然変異率を 0 から大きくしていくと，支配的になる戦略の比率は 1.0 から徐々に小さくなって，0.5 に近づいていく．




(a) 最大多数戦略　(b) 協力率

(c) ALLD　(d) (S-)GRIM　(e) TFT

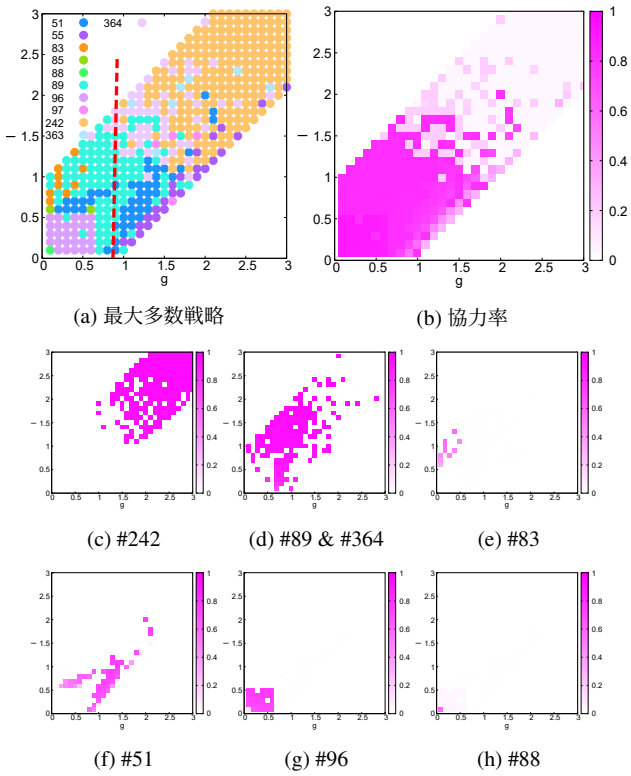
(f) WSLS　(g) FGV　(h) ALLC

図 4: 1 期記憶戦略空間において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 ($\delta = 0.9, p = 0.95, q = 0.01$)

## 5　取り違えがある環境下のダイナミクス

前節では，WSLS と UDMD をゲーム理論的な均衡と 2 戦略のダイナミクスの観点から分析し，これら 2 つの戦略がほぼ無差別であることを明らかにした．本節では，行動の取り違えうる戦略空間を 1 期記憶戦略（32 戦略）からオートマトン戦略の空間（482 戦略）に広げたとき，ダイナミクスの帰結がどう変わるかを分析する．図 4 に 1 期記憶戦略空間の，図 5 にオートマトン戦略空間の結果を示す．ここで，2 人のプレイヤが行動を取り違えない確率 $p = 0.95$，どちらかのプレイヤが行動を取り違える確率 $q = 0.01$ とする．それぞれの図の横軸は自分の裏切りによる利得の増分 $g$，縦軸は相手の裏切りによる損失 $l$ に対応し，0.01 刻みで $[0.01, 3.00]$ をプロットした．また，$|g - l| < 1$ となる組のみを用いた．それぞれの利得パラメータ $g$ および $l$ の組に対して，異なる初期戦略分布をランダムに生成している．図 4a，5a に示した青と赤の点線は，それぞれ式 2 と 3 に $p = 0.95, q = 0.01, \delta = 0.90$ を代入したときの $g$ と $l$ の関係式を示しており，図 4a では WSLS が，図 5a では UDMD がこの線より左側の領域で完全公的均衡となる．

まず，1 期記憶戦略空間におけるダイナミクスの帰結を述べる．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




(a) 最大多数戦略　(b) 協力率

(c) #242　(d) #89 & #364　(e) #83

(f) #51　(g) #96　(h) #88

図 5: オートマトン戦略空間において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 ($\delta = 0.9, p = 0.95, q = 0.01$)

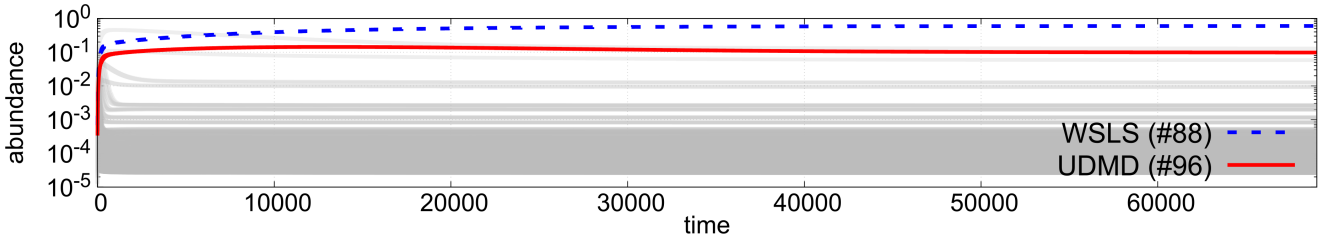
し，協力率は，収束時の戦略分布に対する行動の組 $CC$ の実現頻度である．残りの図 4c-4h は収束時における主要戦略の比率を示している．また，S-GRIM とは，図 1c に示すオートマトンの状態 $P$ から始まる戦略を意味する．図 4d では，GRIM と S-GRIM の比率を合計して 1 つの図とした．

図 4a 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する．まず，$g$ と $l$ が大きいとき，裏切りによる利得の増加，もしくは損失が大きいため，協力を維持しても十分な将来利得を獲得できない．そのため，ALLD (図 1a) が最大多数戦略となり，単独で他戦略を淘汰する．ALLD は相手の行動によらず常に裏切ろうとするため，ALLD が最大多数戦略となるとき，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 $s = 0.03$ がそのまま協力率となる．次に，$g$ および $l$ が中程度のとき，S-GRIM，もしくは GRIM (図 1c) が最大多数戦略となる．GRIM は 2 人のプレイヤが最初はお互いに協力するが，一度でも裏切りが発生すると裏切り続けてしまう．ALLD が最大多数となるときに比べ $g$ や $l$ が小さくなることによりプレイヤは初めは協力を取るようになるが，裏切りの後に協力に戻ろうとはしない．また，1 期記憶戦略における GRIM はどちらかのプレイヤが行動を取り違えるまでは協力状態を維持できる．さらに，2 人のプレイヤが裏切ろうとして同時に取り違えて協力すると協力状態を回復するため，図 4b に示すようにその協力率は 0.690 程度となる．一方で，S-GRIM は初期状態が状態 $P$ となる戦略であり，最初からは協力状態を構成できないため，S-GRIM が最大多数戦略となるときの協力率は 0.169 程度となる．最後に $g$ と $l$ が小さいとき，WSLS (図 1e) が最大多数となる．WSLS 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ため，WSLS の協力率は約 0.9 と非常に高い．

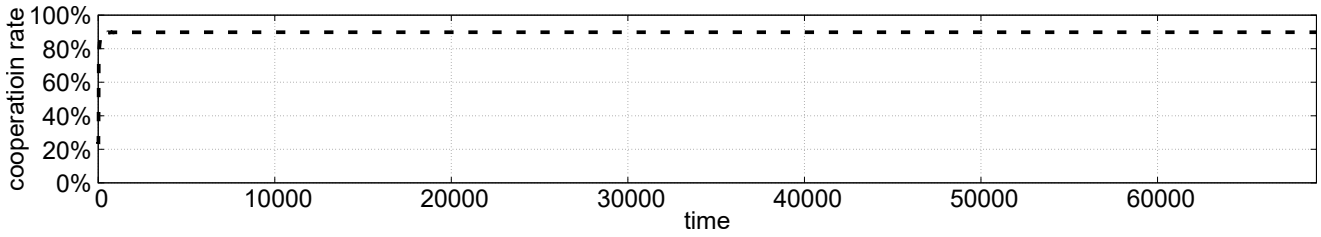
次に，図 5a に示すオートマトン戦略空間の帰結を見ていこう．図 5a に最大多数戦略を，図 5b に協力率を示す．また，残りの図 5c- 5h は収束時における主要戦略の比率を示している．482 戦略の中で，最大多数となった戦略は 11 個あるが，利得パラメータの組において最大多数になる領域を持つ 7 戦略に着目した．1 期記憶戦略空間と同様に，どんな戦略が生き残るかは利得構造に依存し，$g$ および $l$ が小さくなるにつれて，最大多数戦略が協力的になる．$g$ および $l$ が十分大きいと，図 5a に示すように常に裏切ろうとする戦略#242 (図 1a) が他の戦略を支配する．例えば，$g = l = 3.00$ のとき，#242 は 0.967 の比率で生き残り，その協力率は 0.03 となる．$g$ および $l$ がある程度小さくなると，#89 (図 2b) もしくは#364 (図 2d) が最大多数戦略となりやすくなる．これらの戦略は初期状態のみが異なる同じ戦略である．2 人プレイヤがこの戦略にしたがってプレイするとする．このとき，いずれかのプレイヤが行動を取り違えるまでは協力が継続する．いったん取り違えが発生すると，お互いに裏切る状態になるので，2 人が同時に行動を取り違えて協力しない限り協力状態が回復しない．その結果，#89 の協力率は 0.822 と高いが，裏切りから始める#364 のそれは 0.195 にまで下がる．協力率が低くても#364 は生き残りやすいのは，初手に裏切ってくる戦略に対して損を抑えられるためである．これらの戦略に加えて，#51 や#55 といった戦略も最大多数になることもあり，この利得パラメータの領域では，様々な戦略が最大多数になる．しかし一方で，協力率はあまり変化しない．これは，1 つの戦略が他の戦略の淘汰するよりは，複数の戦略が共存したり，サイクルを構成したりすることで，平均的に同じような協力率を達成しているためと考えられる．

最後に，$g$ および $l$ が十分小さくなると，UDMD (#96, 図 2c) もしくは WSLS (#88, 図 2a) が最大多数になりやすくなる．例えば，$g = l = 0.10$ では UDMD および WSLS の比率はそれぞれ 0.098, 0.598 となり，$g = l = 0.50$ では 0.713, 0.045 となった．これはランダムに選ばれた初期分布に依存した結果であるが，図 5g および 5h を比較する限り，1 期記憶戦略との相互作用では最大多数になっていた WSLS を，482 個の戦略との相互作用の結果，UDMD が淘汰していると言える．

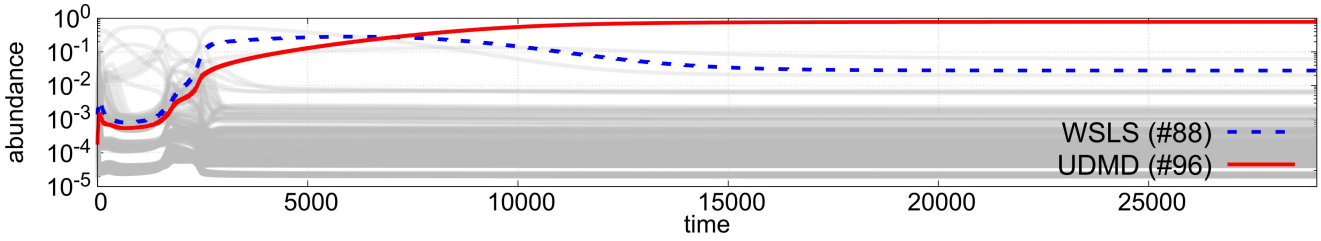
この結果を踏まえて，図 6a および 6b に $g = l = 0.1$ における戦略の比率と協力率の軌跡を示す．このとき，WSLS の初期比率は 0.0007 と，UDMD の 0.0003 より大きいため，WSLS が当初から比率を伸ばし，1000 期で WSLS と UDMD がそれぞれ 0.190 および 0.091 となり，他の戦略より大きな比率を占



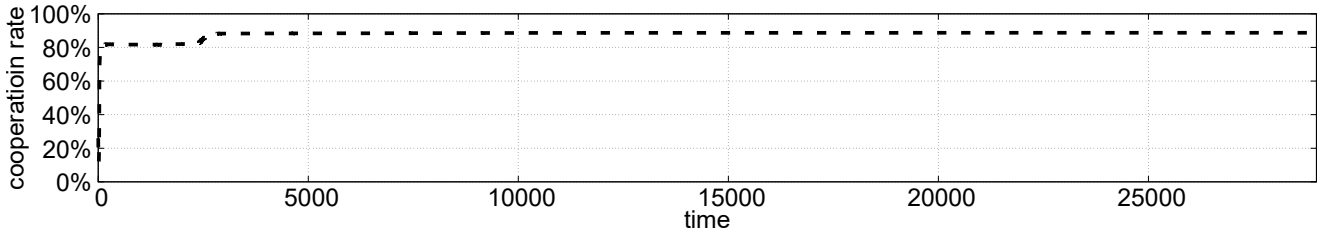

(a) 戦略分布 ($g = 0.10, l = 0.10$)




(b) 協力率 ($g = 0.10, l = 0.10$)




(c) 戦略分布 ($g = 0.50, l = 0.50$)




(d) 協力率 ($g = 0.50, l = 0.50$)

図 6: 戦略分布，および協力率のダイナミクス

めるようになる．その後，WSLS は比率を 0.598 まで伸ばすが，UDMD はそれには届かずに 69000 期ほどで 0.098 に収束した．一方で，協力率は序盤で一気に上昇して 89.8% となり，その後はほぼ変化しなかった．

次に，図 6c および 6d に $g = l = 0.5$ における戦略の比率と協力率の軌跡を示す．このとき，WSLS の初期比率は 0.0014 と，UDMD の 0.0002 であったが，利得パラメータが大きくなり，協力するコストが大きくなるため，最初の 2000 期は多くの戦略が比率を争うことになる．5000 期以降で UDMD と WSLS は最大多数の集団に入ってくるが，その後 WSLS は徐々に比率を下げていく．29000 期ほどで収束し，その時のこれらの戦略の比率は UDMD が 0.782, WSLS が 0.028 となった．一方で，協力率は，序盤で 80% 程度だったが，UDMD が支配的になるにつれて，早い段階で 88.7% で安定し，収束するまで変化しなかった．

## 6 感度分析

本節では，感度分析として，割引因子，行動を取り違えない確率，裏切ったときの利得の増分 (ゲイン) のそれぞれがダイナミクスの帰結に与える影響を吟味する．図 7 に，割引因子 $\delta$ を $[0.800, 0.999]$ の範囲を 0.001 刻みで動かしたときのダイナミクス収束時の戦略の比率を示す．横軸は割引因子 $\delta$ であり，縦軸は戦略の比率を表す．その他のパラメータは $g = l = 0.1$, $p = 0.95, q = 0.01$ とした．図 6 と同様，UDMD を赤，WSLS を青，その他の戦略を灰色とした．このとき，$\delta$ が 0.96 以下の UDMD が，0.96 を超えると WSLS が最大多数となる．割引因子が高ければ高いほど，プレイヤは将来利得を重視するため，裏切られるリスクを負ってでも相互協力を回復させる

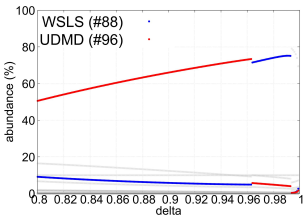



図 7: 割引因子 $\delta$

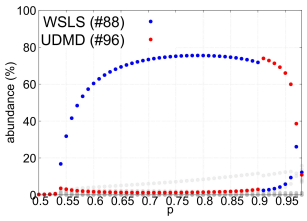
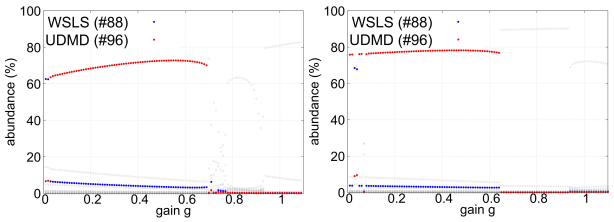



図 8: 取り違えない確率 $p$




(a) $l = 0.1$ のとき　(b) $l = 0.5$ のとき

図 9: ゲイン $g$ に対する収束時の戦略比率

インセンティブが高くなる．その結果，$DD$ だけでなく $CC$ が実現したときでも協力に戻る WSLS にしたがって，相互協力を回復させやすくした戦略が優位になる．

図 8 に，行動を取り違えない確率 $p$ を $[0.50, 0.98]$ の範囲で 0.01 刻みで動かしたときのダイナミクス収束時の戦略の比率を示す．横軸は $p$ であり，縦軸は戦略の比率を表す．図 6 と同様，UDMD を赤，WSLS を青，その他の戦略を灰色で示している．その他のパラメータは $\delta = 0.90, g = l = 0.1, q = 0.01$ とした．このため，$p$ が 0.97 より小さいと，2 人が同時に行動を取り違える確率 $s$ が $q$ より大きくなり，逆に 0.97 より大きいと $s$ が $q$ より小さくなる．このとき，$p$ が 0.90 以下のとき WSLS が，それ以上のとき UDMD が最大多数となる．$p$ が小さくなると行動を取り違えやすくなるため，例えば 2 人のプレイヤが $PP$ にいるとき，意図した $DD$ と異なる行動が実現しやすくなった結果，$CC$ が実現して相互協力を回復させる確率は UDMD より WSLS の方が高くなる．逆に $p$ が大きくなると，意図した $DD$ が実現しやすいので，$DD$ 以外で相互協力を回復させることを考える必要はないため，UDMD が優位になる．

最後に，裏切られたときのロス $l$ を 0.1 もしくは 0.5 に固定して，ゲイン $g$ を変化させたときの収束時の戦略の比率を吟味する．横軸に $g$，縦軸に戦略の比率を表し，その他のパラメータは $\delta = 0.90, p = 0.95, q = 0.01$ とした．図 9a に $l = 0.1$ のときの結果を，図 9b に $l = 0.5$ のときの結果を示す．全体的な傾向はロスの大きさによらず変わらない．ゲイン $g$ がごくごく小さいときだけ WSLS が優位になるが，広い範囲で UDMD が最大多数になっている．実際，$l = 0.1$ のときは $g \leq 0.7$ で，$l = 0.5$ のときは $g \leq 0.65$ で UDMD が最大多数になる．$g$ がこれらの閾値を越えると UDMD も WSLS も淘汰されて#246 や#89, #51 と言った他の戦略が最大多数になる．このことはゲイン $g$ とロス $l$ が大きくなると WSLS も UDMD も均衡を構成しにくくなることと整合している．

## 7　おわりに

本論文では，行動を取り違えうるときの繰り返し囚人のジレンマでどんな戦略が支配的になるかを突然変異付きレプリケータダイナミクスを用いて分析した．従来は 1 期記憶戦略の空間における分析にとどまっており，WSLS がよい性質をもつことが知られていた．これに対して本論文は，今日どの行動をとろうとしたかを含むよう戦略空間を拡張した．その結果，ほぼ全ての 1 期記憶戦略が淘汰される一方で，UDMD という新しい戦略を発見した．さらに，プレイヤの我慢強さと行動を取り違える確率によって WSLS と UDMD のどちらが生き残るかの関係を明らかにした．

## 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] 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.

[4] 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.

[5] 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.

[6] 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.

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

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

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

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

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

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