CF-003

# ほぼ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

五十嵐 瞭平 *　　岩崎 敦 *
Ryohei Igarashi　　Atsushi Iwasaki

## 概要

本論文では，２人のプレイヤがお互いの見間違えをほぼ共有する「ほぼ公的」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．２人がまったく見間違えない「完全」観測下では，常に裏切り (AllD) や 一度でも裏切られたら許さない (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．一方で，ほぼ公的観測下では，有名なしっぺ返し (TFT) が均衡になることが知られている．そこでほぼ公的観測下で,，TFT が生き残るかどうかを検証した．その結果，AllD や GRIM に加えて，相手を 1 度だけ処罰したら，協力に戻る戦略である Forgiver (FGV) が広いパラメータの組で生き残った．一方で，TFT は FGV と比べて狭い範囲でしか生き残らない．正確には，お互いの見間違えを共有するときは FGV が他の戦略を淘汰しやすい．一方で徐々に見間違えが共有されにくくなると，TFT と AllC の混合戦略が生き残りやすくなることを示した．

## 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であり [1, 2]，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, 4]．本論文では「ほぼ公的」観測下の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を用いて分析した．2 人がまったく見間違えない「完全」観測下では，常に裏切り（AllD）や一度でも裏切られたら許さない（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 [5]．意外なことに有名なしっぺ返し（Tit-For-Tat, TFT）が生き残ることはない．

一方で，2 人のプレイヤがお互いの見間違えをほぼ共有する「ほぼ公的」観測下では TFT が均衡になることが知られている [6]．そこで，ほぼ公的観測下で TFT が生き残るかどうか知ることが本論文の最初の動機づけとなる．繰り返しゲームの戦略は，今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 以下の有限状態機械 (Finite State Automaton，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 [7] 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 [8, 9]．その結果，ほぼ公的観測下では AllD や GRIM に加えて，相手を 1 度だけ処罰したら協力に戻る戦略である Forgiver（FGV）[10] が幅広い利得パラメータの組において生き残ることがわかった．一方で，TFT は FGV と比べて狭いパラメータの組でしか生き残らなかった．TFT は相手の行動を真似るため，どんな戦略と対戦しても大きく損をしない構造をもっているものの，ほぼ公的観測下でも協力状態を回復しにくいため，レプリケータダイナミクスで生き残りにくかったと考えられる．一方で，相手をすぐに許す FGV の方が，協力状態を回復しやすく，ダイナミクスの過程で生き残る戦略と協力状態を維持しやすいことを明らかにした．

## 2　モデル

本章では文献 [11] に基づいて，2 人私的観測付き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 $i$ は有限集合 $A$ 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．次に，プレイヤ $i$ は $\boldsymbol{a}$ に関する私的なシグナル $\omega_i \in \Omega$ 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 $o(\boldsymbol{w} \mid \boldsymbol{a})$ 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繰り返し行われるので，プレイヤ $i$ の割引利得和は割引因子 $\delta \in (0, 1)$ により $\sum_{t=1}^{\infty} \delta^t g_i(\boldsymbol{a}^t)$ となる. ただし，$g_i(\cdot)$ の値は利得表によって定められた値に従う．

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，囚人にジレンマでは $|g - l| < 1$ が要求される．もしこの条件が成り立たないとすると，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．

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

* 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 $|g - l| < 1$)

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

表 2: $(a_1, a_2)$ のときのシグナル分布

| | $w_2 = g$ | $w_2 = b$ |
|---|---|---|
| $w_1 = g$ | $P_{a_1,a_2}(1-e)^2 + (1 - Pa_1, a_2)e^2$ | $e(1-e)$ |
| $w_1 = b$ | $e(1-e)$ | $P_{a_1,a_2}e^2 + (1 - Pa_1, a_2)(1-e)^2$ |

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．

ここで，ほぼ公的観測のシグナル分布を定義する．プレイヤ 1 と 2 の行動プロファイルが $(a_1, a_2)$ のとき，公的シグナルが $GOOD$ となる確率を $P_{a_1,a_2}$，$BAD$ となる確率を $1 - P_{a_1,a_2}$ とする．ほぼ公的観測では公的シグナルを一定の確率 $e$ で見間違える．そのため，例えば公的シグナルが $GOOD$ であるとき両プレイヤは $(1-e)^2$ の確率で正しい観測をし，シグナルプロファイルは $(g, g)$ となる．反対に $e^2$ で両プレイヤはともに見間違え，シグナルプロファイルは $(b, b)$ となる，また，一方のプレイヤのみの見間違えは $e(1-e)$ で発生し，シグナルプロファイルは $(g, b)$ または $(b, g)$ となる．公的シグナルが $BAD$ であるときも同様に，シグナルプロファイルを与える確率を計算でき，ほぼ公的観測のシグナル分布は表 2 に従う．

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．有限状態機械 (Finite State Automaton，FSA) は繰り返しゲームの戦略を簡略に表記する方法であり，本研究では，状態数 2 以下の非同相な 26 個の FSA を用いる．

このような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た突然変異付きレプリケータダイナミクスを用いる．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ようになる．すなわち，ある戦略が突然変異する確率を $u$ とおくと，突然変異付きレプリケータ方程式は

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

と定義する [10]．$\phi(\cdot)$ 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る．

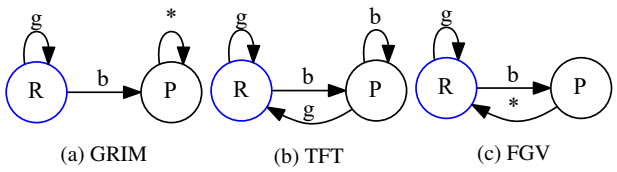



図 1: 主要な FSA

数値実験では，割引利得 ($\delta = 0.90$) を固定した上で，$g, l$ を $[0.01, 1.00]$ の範囲で 0.01 刻みで変化させた．戦略として状態数 2 以下の FSA26 個を用いる．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，各戦略の存在比率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．ダイナミクスは全ての戦略の 1 期あたりの人口の変化量 $|\dot{x}|$ が $10^{-5}$ 以下となった時点で収束と判定した．また，1000 期までに収束と判定されなかった場合は計算を終了する．

## 3 主要 6 戦略とナッシュ均衡

本節では，繰り返し囚人のジレンマにおいて主要な 6 戦略とそのナッシュ均衡を概説する．繰り返しゲームの戦略は過去の行動と観測の履歴から現在の行動への写像で定義される．一般には複雑になる戦略でも FSA を用いることで簡略に表記できる．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．状態数 1 の戦略には AllC と AllD の 2 つが存在し，AllC は状態 R のみを持ち毎期必ず協力する戦略，AllD は状態 P のみを持ち毎期必ず裏切る戦略である．

本研究では，状態数 2 以下の非同相な 26 個の FSA からなる戦略空間をもつゲームを考える [13]．26 戦略の中から，完全観測でお互いの協調を均衡で維持する 5 つの戦略に AllC を加えたものを主要 6 戦略と定義する．まず無限期罰則のトリガー戦略 (Grim Trigger，GRIM) を説明する (図 1a)．GRIM は最初に協力し，相手の裏切りを観測するとそれ以降裏切り続ける戦略である．主要 6 戦略でもっとよく知られているのは “しっぺ返し” (Tit-For-Tat，TFT) であろう (図 1b)．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では協力を，裏切りを観測した次の期には裏切りを行う戦略である．

さらに本論文がその重要性を明らかにした戦略として，“Forgiver”（FGV）を説明する (図 1c) [10]．FGV は状態 R からスタートし，相手の裏切りを観測したら状態 P に遷移して，次の期に裏切る．その後は，どちらのシグナルを観測しても状態 R に遷移して協力に戻る．つまり，一度だけ相手を処罰したら必ず協力へ戻る寛容な戦略である．最後に，“勝ち残り，負け逃げ” (Win-Stay，Lose-Shift, WSLS) [14] があるが，これ

表 3: 主要 6 戦略を抜粋した利得表 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01, g = 0.10, l = 0.10$)

| Strategy | AllC | AllD | FGV | TFT | WSLS | GRIM |
|---|---|---|---|---|---|---|
| AllC | 1.000 | −0.100 | 0.903 | 0.809 | 0.855 | 0.458 |
| AllD | 1.100 | 0.000* | 0.715 | 0.451 | 0.632 | 0.172 |
| FGV | 1.009 | −0.065 | 0.905* | 0.794 | 0.842 | 0.468 |
| TFT | 1.017 | −0.041 | 0.885 | 0.787 | 0.818 | 0.482 |
| WSLS | 1.013 | −0.057 | 0.881 | 0.752 | 0.840 | 0.467 |
| GRIM | 1.049 | −0.016 | 0.821 | 0.664 | 0.768 | 0.492* |

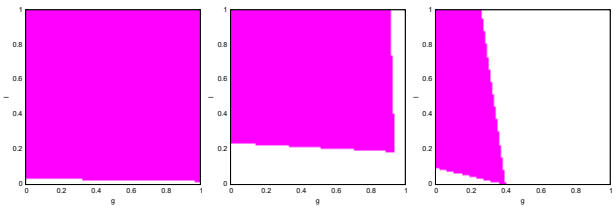
表 4: 主要 6 戦略を抜粋した利得表 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.03, g = 0.10, l = 0.10$)

| Strategy | AllC | AllD | FGV | TFT | WSLS | GRIM |
|---|---|---|---|---|---|---|
| AllC | 1.000 | −0.100 | 0.890 | 0.787 | 0.835 | 0.420 |
| AllD | 1.100 | 0.000* | 0.717 | 0.458 | 0.631 | 0.173 |
| FGV | 1.010 | −0.065 | 0.882 | 0.764 | 0.808 | 0.417 |
| TFT | 1.019 | −0.042 | 0.867 | 0.755 | 0.792 | 0.432 |
| WSLS | 1.015 | −0.057 | 0.857 | 0.721 | 0.788 | 0.412 |
| GRIM | 1.053 | −0.016 | 0.805 | 0.636 | 0.743 | 0.438* |

は本論文の範囲では生き残ることはない.

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 の組がナッシュ均衡になるとは，その組がお互いに最適反応 FSA になっていることを言う [15]．表 3 および 4 に 26 戦略間から主要 6 戦略を抜粋した平均利得（正規化した割引利得和）の表を示す．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 26 戦略間でのナッシュ均衡になっていることを表す．表 3 のシグナル分布と利得のパラメータはそれぞれ $P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01, g = 0.10, l = 0.10$ である．ここでは，AllD，GRIM および FGV がそれぞれ均衡を構成している．主要 6 戦略以外では，最初に状態 P から始める FGV が均衡を構成する．見間違えを共有する度合いを小さくした (独立誤差 $e$ を 0.03 に増加させた) 表 4 では，FGV が均衡を構成しなくなり，AllD と GRIM のみが均衡を構成する．また主要 6 戦略以外で均衡を構成する戦略は存在しない.

図 2 に GRIM，TFT および FGV がナッシュ均衡となる利得パラメータの範囲を示す．GRIM は均衡を形成しやすい戦略であり，裏切られることで被る損失 $l$ が非常に小さい時を除いて均衡を構成する．しかし，表 3 が示すように GRIM 同士の対戦が実現する利得は FGV や TFT に比べて小さい. TFT は裏切ることで得る利得の増加 $g$ が 0.9 より小さく，$l$ が 0.2 より大きいとき，均衡を構成する．一方で，FGV が均衡を構成する $g$ の上限は 0.4 と TFT より低いが，$l$ が 0.2 より小さくても均衡を構成することがある.



(a) GRIM (b) TFT (c) FGV

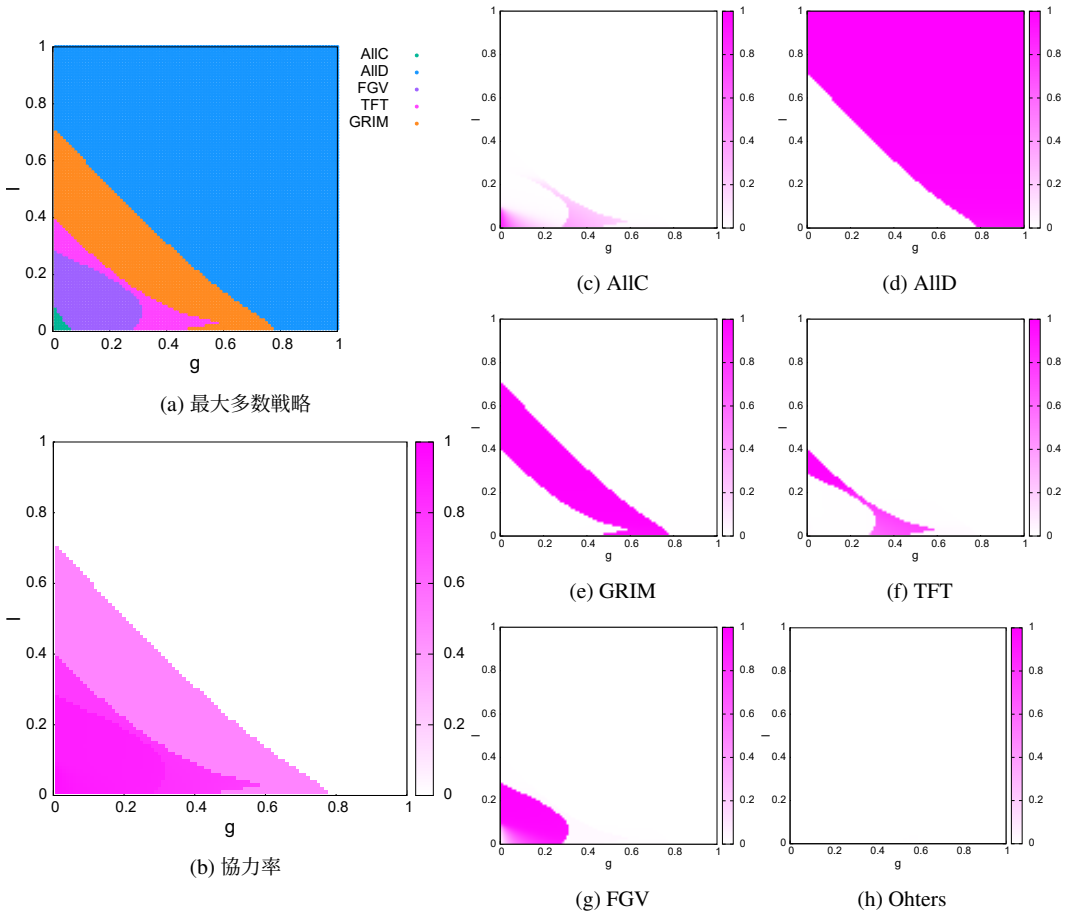
図 2: GRIM,TFT および FGV が均衡となる利得パラメータ範囲 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01$)

## 4 ほぼ公的観測下のダイナミクス

図 3 にほぼ公的観測下におけるダイナミクスの帰結を示す．ここで，シグナル分布パラメータは $P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01$ 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 $[0.01, 1.00]$ をプロットした．図 3a に最大多数戦略を，図 3b 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，協力率は，収束時の戦略分布に対する $(C, C)$ の実現頻度である．また，残りの図 3c-3h は収束時における主要戦略の割合を示している.

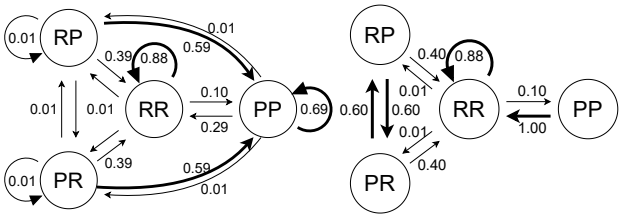
図 3a が示すように，どんな戦略が生き残るかは利得構造に依存する. まず，$g$ と $l$ が大きいとき，裏切る誘引や裏切られることによる損失が大きいため，他のどの戦略も協力を維持するに十分な将来利得を獲得できない．そのため，AllD が最大多数戦略となり，単独で他戦略を淘汰する. 次に，$g$ および $l$ が中程度のとき，GRIM が最大多数戦略となる. このとき，プレイヤは最初はお互いに協力するが，一度でも裏切りが発生すると永遠に裏切り続けてしまう. AllD が最大多数となるときに比べ $g$ や $l$ が小さくなることによりプレイヤは初めは協力を取るようになるが，裏切りの後に協力に戻ろうとはしない. また，GRIM は見間違えが起こるまでは協力状態を維持できるため，図 3b に示すようにその協力率は 0.500 程度となる.

最後に $g$ と $l$ が小さいとき，FGV が広い範囲で最大多数となる．またこの領域において TFT も最大多数となるが，その利得パラメータの範囲は FGV に比べて狭く，GRIM と FGV に挟まれたわずかな領域のみである. 見間違えを共有するのであれば，見間違えが発生したときプレイヤはどちらも $bad$ を観測している．このとき，TFT 同士の対戦を考えると互いに $bad$ を観測しているため，次の期では裏切り合いが発生する. 図 4a が示すように，裏切り合いの状態である状態 PP から最も起きやすい遷移は状態 PP への遷移であり，裏切り合いから簡単に抜け出すことはできない. この裏切り合いの状態から再び協力状態に戻るためには，確率 0.29 で実現する状態 PP から状態 RR へ遷移を待つ必要がある. しかし，FGV 同士の対戦であれば，TFT 同士の対戦と同様に裏切り合いが発生したとしても図 4b が示すように状態 PP から状態 RR へ確実に戻る




(a) 最大多数戦略　(b) 協力率　(c) AllC　(d) AllD　(e) GRIM　(f) TFT　(g) FGV　(h) Ohters

図 3: ほぼ公的観測下のダイナミク ($P_{cc} = 0.90, P_{cd} = P_{dc} = 0.40, P_{DD} = 0.30, e = 0.01$)




(a) TFT　(b) FGV

図 4: 積 FSA ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01$)

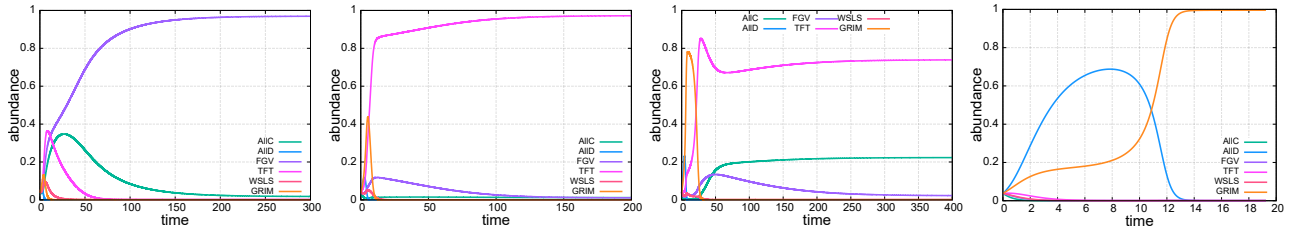
ことが出来るため，簡単に協力状態を回復できる．FGV のこの機構が TFT を抑えて，広いパラメータ範囲で最大多数戦略となる秘訣である．また，この領域における協力率は $g$ や $l$ の値により変化し，その範囲はおおよそ 0.70 から 0.90 となる．

図 5 に 4 つの利得パラメータ $(g, l) \in \{(0.10, 0.10), (0.10, 0.25), (0.35, 0.10), (0.35, 0.25)\}$ における戦略比率のダイナミクスを表す．図 5a は FGV が，図 5b は TFT が他の戦略を 100 期程度でほぼ淘汰している．図 5c では，TFT, AllC, FGV の 3 つの戦略が共存する状態から，FGV が淘汰され，TFT と AllC が 0.739 : 0.223 の割合で共存するようになる．最後に図 5d は GRIM がたった 20 期ほどで他の戦略を淘汰する．

## 5 議論

### 5.1 独立誤差 $e$ の影響

ほぼ公的観測はプレイヤは見間違えをほぼ共有するという特徴を持つが，どの程度見間違えを共有するかは独立誤差 $e$ によって決まる．$e$ が小さければ見間違えは共有されるが，逆に大きければ見間違えは共有されにくくなり，プレイヤが異なるシグナルを観測しやすくなる．前章において FGV の生存には見間違えが共有されていることが重要であることを示した．そこで，見間違えの共有度合いが FGV の生存にどのような影響を与えるかを調べた．図 6 に $g = 0.10, l = 0.10$ において，$e$ を変化させ収束時の戦略の比率を示した．見間違えが共有されやすい ($e = 0.01$) ときは，FGV が他の戦略をほぼ淘汰する．一方で，$e$ が増加して，見間違いが共有されにくくなると，FGV はその比率を徐々に減らしていく．そして，$e$ が 0.026 を越えると FGV は瞬間的にほぼ生き残らなり，代わりに TFT と AllC が共存するようになる．




(a) $g = 0.10, l = 0.10$　(b) $g = 0.10, l = 0.25$　(c) $g = 0.35, l = 0.10$　(d) $g = 0.35, l = 0.25$

図 5: 4 つの利得パラメータにおける戦略比率のダイナミクス ($P_{CC} = 0.90$, $P_{CD} = P_{DC} = 0.40$, $P_{DD} = 0.30$, $e = 0.01$)

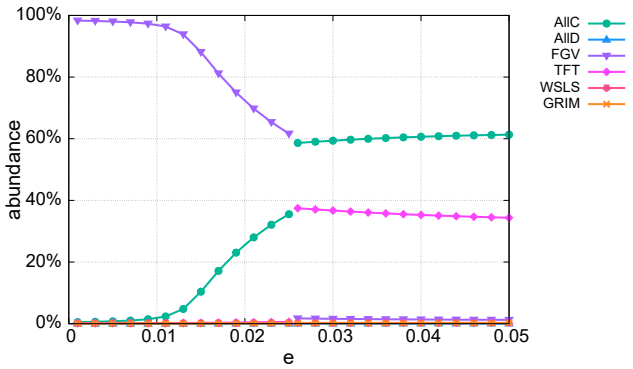



図 6: 独立誤差 $e$ による戦略比率の変化 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, g = 0.10, l = 0.10$)

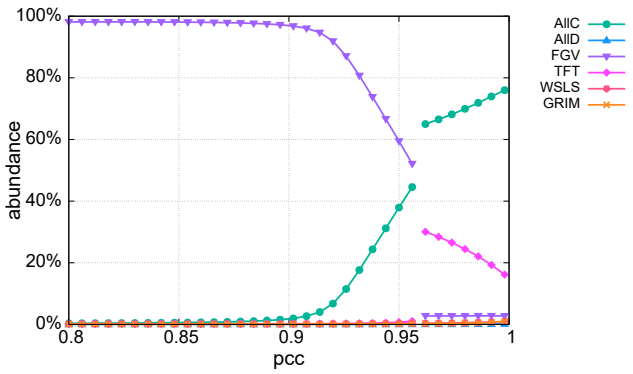
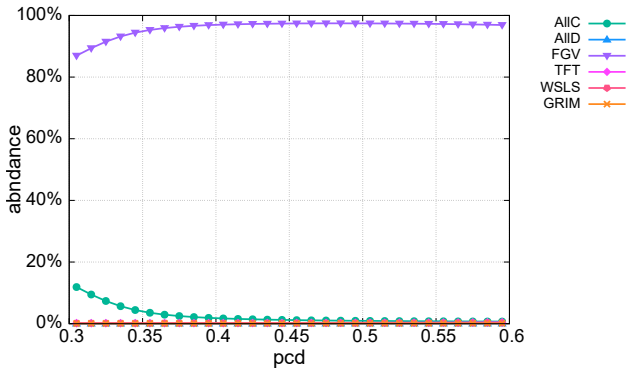
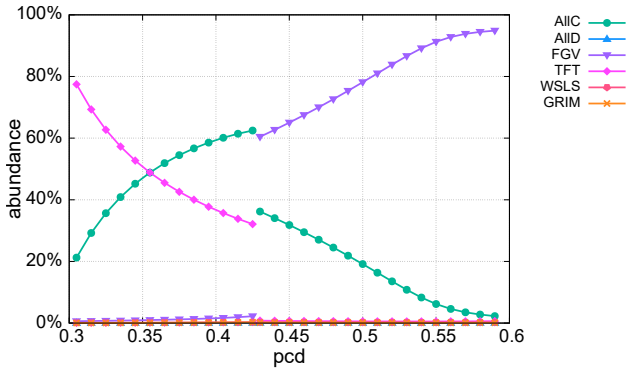



図 7: シグナルパラメータ $P_{CC}$ による戦略比率の変化 ($P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01, g = 0.10, l = 0.10$)




(a) $e = 0.01$



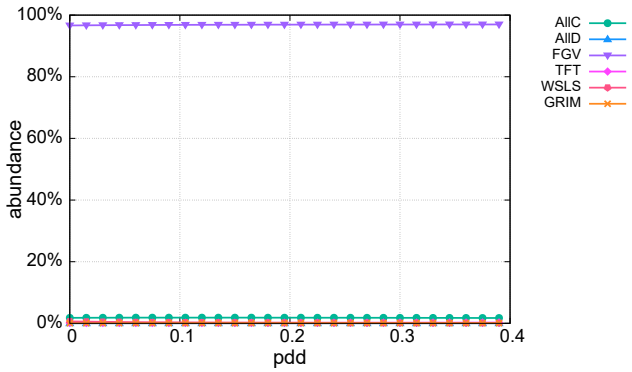

(b) $e = 0.03$

図 8: シグナルパラメータ $P_{CD}$ による戦略比率の変化 ($P_{CC} = 0.90, P_{DD} = 0.30, e = 0.01, g = 0.10, l = 0.10$)

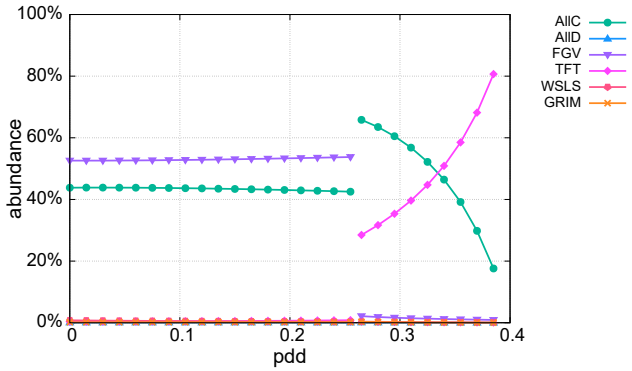
ここで，表 4 に示すように $e = 0.03$ では，FGV だけでなく，TFT や AllC も均衡を構成していない．実はここで TFT と AllC は $(0.378, 0.622)$ の比率で混合戦略均衡を構成している．実際，この共存状態では，(TFT,AllC)=$(0.367, 0.594)$ となっており，上記の混合戦略均衡とほぼ一致する．

### 5.2　公的シグナルのパラメータの影響

本節では公的シグナルのパラメータがダイナミクスの帰結に与える影響を分析する．図 7 に相互協力時に公的シグナルが $GOOD$ となる確率である $P_{CC}$ に対する収束時の戦略比率の変化を表す．具体的には，$P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01, g = 0.10, l = 0.10$ に対して，$P_{CC}$ を $[0.80, 0.99]$ の範囲で 0.001 刻みで変化させたときの主要 6 戦略の比率を示す．$P_{CC}$ が小さいときは，FGV が最大多数となり他の戦略をほぼ支配する．$P_{CC}$ が増加するにつれて FGV は比率を徐々に減らし，AllC の比率が増加する．$P_{CC}$ が 0.96 を越えた瞬間に，FGV は一気に淘汰され，代わりに TFT と AllC が共存するようになる．以上から，相互協力時の公的シグナルが十分に正確であれば，TFT と AllC の混合戦略が生き残るようになる．さらに図 6 で示したように TFT と AllC の混合戦略は独立誤差が大きいときにも生き残りやすい．逆に言えば，FGV は相互協力時の公的シグナルがある程度不正確だが，プレイヤ間での見間違えが起こりにくいときに生き




(a) $e = 0.01$




(b) $e = 0.03$

図 9: シグナルパラメータ $P_{DD}$ による戦略比率の変化 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, e = 0.01, g = 0.10, l = 0.10$)

残りやすくなると言える．

次に 2 人のうちいずれかのプレイヤが裏切るときの公的シグナルが $GOOD$ になる確率 $P_{CD}$ ($P_{DC}$) を変化させて，戦略比率の変化を観察する．図 8 に $g = 0.10, l = 0.10$ に対して，$P_{CD}$ を $[0.30, 0.60]$ の範囲で 0.001 刻みで変化させたときの主要 6 戦略の比率を示す．まず図 8a は，独立誤差が小さい ($e = 0.01$) とき，FGV が，$P_{CD}$ の値によらず，常に最大多数となり他の戦略をほぼ支配することを示している．次に図 8b は，独立誤差が大きい ($e = 0.03$) とき，TFT と AllC の混合戦略と FGV のいずれかが最大多数となることを示している．つまり，$P_{CD}$ が小さいとき，TFT と AllC が共存して他戦略を支配する．$P_{CD}$ が大きくなるにつれて，AllC の比率が増加し，その値が 0.36 を越えたとき，最大多数戦略が TFT から AllC に代わる．そして，$P_{CD}$ が 0.42 を越えるのを境に TFT と AllC の共存は見られなくなり，FGV と AllC の共存が発生するようになる．$P_{CD}$ の増加とともに FGV の比率は増加し，0.60 以上では FGV の比率がほぼ 1.0 になる．

$P_{CD}$ はいずれかのプレイヤが裏切ったときに公的シグナルが $GOOD$ となる確率である．すなわち，$P_{CD}$ の値は 1 人の裏切りが全体の結果に与える影響の大きさを決めることになる．このため，$P_{CD}$ が小さいときは，1 人が裏切るとすぐに $BAD$ が出やすくなるので，1 人の裏切りが全体の結果に与える影響が大きいといえる．このようなとき，図 8b は TFT と AllC のいずれかを確率的に選ぶことを推奨しているといえる．逆に，1 人の裏切りの影響が小さいときは，FGV と AllC のいずれかを確率的に選ぶことを推奨している．ただし，$P_{CD}$ がこのような影響を与えるのは，独立誤差 $e$ がある程度大きいときであるので，独立誤差 $e$ が十分に小さく，お互いの見間違えを共有しやすいときは，$P_{CD}$ の大きさによらず，FGV にしたがって振る舞うことが推奨される．

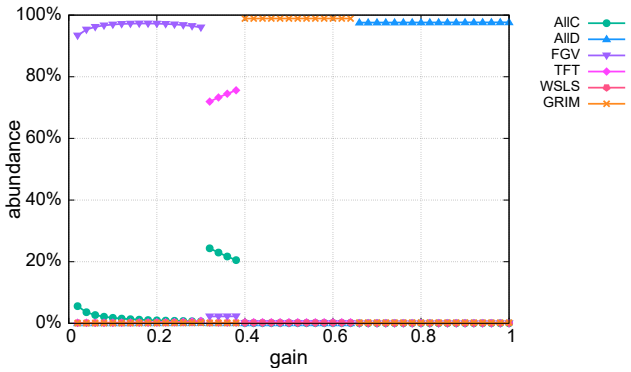
最後にプレイヤがお互いに裏切るとき，公的シグナルが $GOOD$ になる確率 $P_{DD}$ を変化させて，戦略比率の変化を観察する．図 9 に $P_{DD}$ を $[0.00, 0.40]$ の範囲で 0.001 刻みで変化させたときの主要 6 戦略の比率を示す．まず図 9a は，独立誤差が小さい ($e = 0.01$) とき，FGV が，$P_{DD}$ の値によらず，常に最大多数となり他の戦略をほぼ支配することを示している．

次に図 9b は，独立誤差が大きい ($e = 0.03$) とき，FGV と AllC の混合戦略と TFT と AllC の混合戦略のいずれかが最大多数となることを示している．つまり，$P_{DD}$ が小さいとき，FGV と AllC が共存して他戦略を支配する．$P_{DD}$ が大きくなっても両者の比率はそれほど変化しないが，その値が 0.26 を越えたとき，TFT と AllC の混合戦略にとって代わられる．そして，$P_{DD}$ の増加とともに TFT の比率が増加し，0.34 以上で TFT が最大多数戦略となる．$P_{DD}$ はプレイヤがお互いに裏切ったときに公的シグナルが $GOOD$ となる確率である．すなわち，$P_{DD}$ の値が小さいほど公的シグナルは正確だといえる．このため，お互いの裏切りが正確にわかるときは，FGV か AllC をのいずれかを確率的に選ぶことが推奨される．逆に徐々に正確さを減るにつれて，TFT か AllC をのいずれかを確率的に選ぶことが推奨されるようになる．ただし，$P_{DD}$ がこのような影響を与えるのは，独立誤差 $e$ がある程度大きいときであるので，独立誤差 $e$ が十分に小さく，お互いの見間違えを共有しやすいときは，$P_{CD}$ のときと同様，$P_{DD}$ の大きさによらず，FGV にしたがって振る舞うことが推奨される．

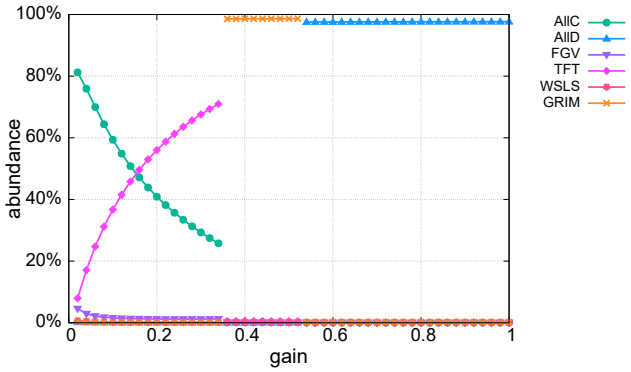
### 5.3 裏切りによる利得の増分の影響

本節では，裏切ることで得られる利得の増分（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．図 10a に，裏切られたときの損失 $l$ を 0.10 に，独立誤差 $e$ を 0.01 に固定したときの結果を示す．$g$ が 0.31 より小さいときは FGV が最大多数となり，0.31 を越えると，TFT と AllC の混合戦略が最大多数となる．さらに，0.38 を越えると GRIM が，0.66 を越えると AllD が最大多数となる．

図 10b に独立誤差 $e$ を 0.03 に変えたときの結果を示す．$g$ が 0.36 より小さいときは TFT と AllC の混合戦略が最大多数となり，0.36 を越えると GRIM が最大多数に，さらに 0.54 を越えると AllD が最大多数となる．TFT と AllC の混合戦略が最大多数となるとき，どちらの戦略が多くなるかは $g$ に依存する．具体的には $g$ が 0.18 より小さいときは AllC が，大きいときは TFT が最大多数となっている．

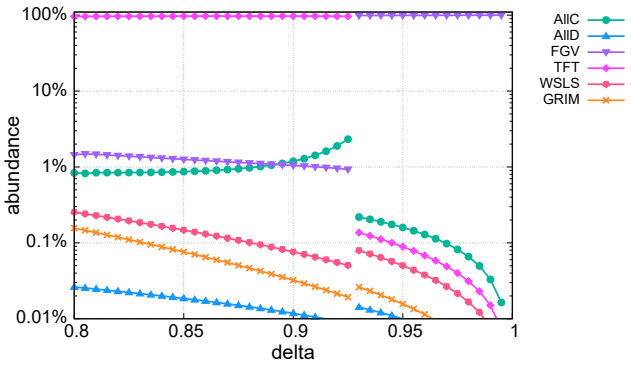



(a) $e = 0.01$

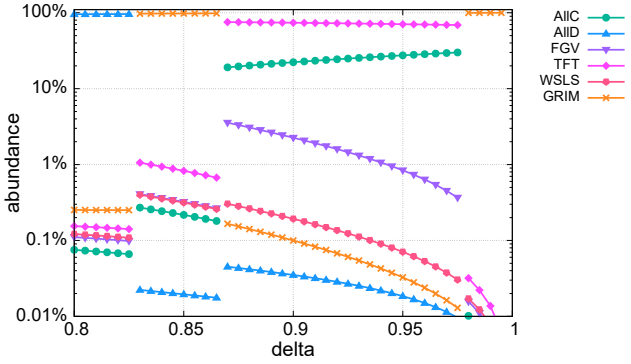



(b) $e = 0.03$

図 10: 裏切りで得る利得の増分 $g$ に対する戦略人口と期待利得 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, l = 0.1$)




(a) $g = 0.10, l = 0.25$




(b) $g = 0.35, l = 0.10$

図 11: 割引因子 $\delta$ による戦略比率の変化 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01$)

### 5.4 割引因子の影響

本節では，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 11 に割引因子 $\delta$ を $[0.80, 0.99]$ の範囲を 0.001 刻みで動かしたときの主要 6 戦略の戦略比率の変化を示す．ここまでの分析で主に使っていた利得パラメータの組 $g = 0.10, l = 0.10$ では，割引因子の値によらず FGV が最大多数となる．そこで，TFT が最大多数となる 2 つの利得パラメータの組に着目する．図 11a は $g = 0.10, l = 0.25$ の，図 11b は $g = 0.35, l = 0.10$ の結果を表す．これらの図の横軸は割引因子 $\delta$ であり，縦軸は戦略の比率を表す．ただし，縦軸は対数スケールになっている．

まず，図 11a では，$\delta$ が十分小さいと TFT が他の戦略をほぼ淘汰する．$\delta$ が増加しても傾向は変わらないが，0.93 を越えると TFT に代わり，FGV が他の戦略を淘汰するようになる．このため，裏切ることで得る利得の増分 $g$ が裏切られたときの損失 $l$ よりも小さい範囲では，プレイヤが我慢強くなることで FGV が他の戦略を支配する範囲が拡大すると言える．

次に，図 11b では，$\delta$ が十分小さいと AllD，そして GRIM が他の戦略をほぼ淘汰するようになる．さらに $\delta$ が増加し，0.87 を越えると，GRIM に代わって TFT と AllC の混合戦略が他の戦略を淘汰するようになる．そして，0.98 を越えると GRIM が再び最大多数戦略になる．いっけん，プレイヤが我慢強くなり，将来の利得を重要視するようになると，FGV や TFT と AllC の混合戦略のように見間違えの後でも協調状態に戻りやすい戦略が有利になると考えられる．しかし，$g$ が $l$ よりも大きい範囲では，割引因子が大きすぎても GRIM のような不寛容な戦略しか生き残らなくなる．

### 5.5 初期戦略分布とダイナミクスの帰結の関係

本節では，ダイナミクスに与える初期の戦略分布 (初期戦略分布) が与える影響を吟味する．ここまで初期戦略分布は一様である，つまり各戦略は 1/26 ずつの比率をからダイナミクスを始めるとしていた．レプリケータダイナミクスの帰結は初期戦略分布によって変化するため，得られた結果が初期戦略分布を一様分布としたときに限られるのか否かを吟味する．そこで初期戦略分布をランダムに 10000 個を生成して，それぞれの帰結を計算し，どのような戦略の比率に収束したかを分類した．表 5 は独立誤差 $e$ が 0.01 としたとき，4 つの利得パラメータの組での結果を表す．例えば，$g = 0.1, l = 0.1$ のとき，AllD が最大多数となった割合は 0.02, GRIM が最大多

表 5: 代表点における帰結の実現割合 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.01$)

| | (g,l) | | | |
|---|---|---|---|---|
| RMD の帰結 | $(0.10, 0.10)$ | $(0.10, 0.25)$ | $(0.35, 0.10)$ | $(0.35, 0.25)$ |
| AllD only | 0.01 | 0.10 | 0.18 | 0.36 |
| GRIM only | 0.01 | 0.26 | 0.31 | $0.62^{\dagger}$ |
| FGV only | $0.90^{\dagger}$ | 0.25 | 0.09 | - |
| TFT only | - | $0.38^{\dagger}$ | - | 0.02 |
| TFT-AllC | 0.07 | - | $0.41^{\dagger}$ | - |
| FGV-AllC | - | - | - | - |
| Others | 0.01 | 0.01 | 0.01 | - |

表 6: 代表点における帰結の実現割合 ($P_{CC} = 0.90, P_{CD} = P_{DC} = 0.40, P_{DD} = 0.30, e = 0.03$)

| | (g,l) | | | |
|---|---|---|---|---|
| RMD の帰結 | $(0.10, 0.10)$ | $(0.10, 0.25)$ | $(0.35, 0.10)$ | $(0.35, 0.25)$ |
| AllD only | 0.02 | 0.18 | 0.28 | 0.50 |
| GRIM only | 0.01 | 0.35 | 0.32 | $0.50^{\dagger}$ |
| FGV only | - | - | - | - |
| TFT only | - | - | - | - |
| TFT-AllC | $0.67^{\dagger}$ | $0.46^{\dagger}$ | $0.39^{\dagger}$ | - |
| FGV-AllC | 0.29 | - | - | - |
| Others | 0.01 | 0.02 | 0.01 | - |

数となった割合は 0.01, FGV が最大多数となった割合は 0.9, TFT と AllC の混合戦略となった割合は 0.07 となった．ここで FGV の割合 0.9 に † をつけているが，これは初期戦略分布が一様であるときの帰結を意味する．どの利得パラメータを見ても初期戦略分布が一様であるときの帰結に到達する割合がもっとも高いことがわかる．このため，見間違えが共有されやすい，かつ $g$ と $l$ が十分小さいときは，初期戦略分布に対しても FGV が頑健であることがわかる．

表 6 は独立誤差 $e$ が 0.03 としたときの結果を表す．$e = 0.01$ のときと異なり，FGV や TFT が単独で他戦略を淘汰するようなことが全く実現していない．一方で，TFT と AllC や，FGV と AllC の混合戦略が現れるようになっている．とくに，TFT と AllC の混合戦略は $g = 0.35, l = 0.25$ を除くケースで 4 割から 6 割強の割合で到達している．したがって，見間違えが共有されにくくなるときは，どのような初期戦略分布であっても，TFT もしくは FGV にしたがって振る舞いつつ，たまに AllC に従うことで高い利得が実現できる．

## 6 おわりに

本論文では，２人のプレイヤがお互いの見間違えをほぼ共有する，ほぼ公的観測下の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を用いて分析した．TFT はほぼ公的観測下で均衡になることが知られていたが，その戦略が単独で生き残るパラメータの範囲はかなり狭いことがわかった．代わりに相手を一度処罰したらすぐに許す FGV という戦略が生き残りやすくなることを世界で初めて明らかにした．とくにお互いの見間違えを共有するときは，FGV が他の戦略をしばしば淘汰する．また，見間違えが共有されにくくなると，TFT と AllC の混合戦略，つまり確率的に常に協力としっぺ返すのを混ぜることで高い利得を維持できることがわかった．

## 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] 岡田章. ゲーム理論 新版. 有斐閣, 2011.

[5] 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. 第 19 回情報科学技術フォーラム, 2020.

[6] Christopher Phelan and Andrzej Skrzypacz. Beliefs and Private Monitoring. *Review of Economic Studies*, Vol. 79, No. 4, pp. 1637–1660, 2012.

[7] Lorens 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.

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

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

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

[11] ジョヨンジュン, 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り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, pp. 1234–1246, 2012.

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

[13] Martin A. Nowak, Karl Sigmund, and Esam El-Sedy. Automata, repeated games and noise. *Journal of Mathematical Biology*, Vol. 33, pp. 703–722, 1995.

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

[15] 神取道宏. ミクロ経済学の力. 日本評論社, 2014.