

# 異なる利得構造を持つエージェントが混在するネットワーク における協調促進について

村中 慧<sup>a)</sup> 大塚 知亮<sup>b)</sup> 菅原 俊治<sup>c)</sup>

概要：本研究では、異なる利得行列を持つエージェントが混在するネットワークにおいて、これまで著者らが提案した強化協調期待戦略の協調促進の頑健性を調査する。囚人のジレンマでは、裏切り行動がナッシュ均衡であり、合理的な行動を取ると裏切り行動を選択することになる。しかし、パレート最適な行動は協調であり、社会全体の利益を増大させるためには、協調を促進することが重要である。これに対し、複雑ネットワーク全体に協調を促進させる強化協調期待戦略を提案した研究がある。先行研究では、すべてのエージェントが囚人のジレンマゲームの利得行列を持つとしていたが、実社会ではエージェントの行動に関する価値観が同一とは限らない。そのため、先行研究では、常に協調しないあるいは確率的に協調しないエージェントを混在させ、協調促進の頑健性を示していた。本研究では、その他の戦略として、一部のエージェントの利得構造を反調整ゲームとするものを混在させたネットワーク環境での頑健性を調査する。

MURANAKA KEI<sup>a)</sup> OTSUKA TOMOAKI<sup>b)</sup> SUGAWARA TOSHIHARU<sup>c)</sup>

## 1. 序論

近年のコンピュータの発展により、PC やスマートフォンなどの様々な端末にコンピュータシステムが搭載されるようになった。こうしたデバイス上で動作するエージェントは、ユーザに代わって情報を収集、提供し、ユーザ間の交渉や取引を行う。このとき、エージェント間のインタラクションの際に、ユーザにとって適切な振る舞いをし、適切な行動戦略を取ることが必要である。そのため、システムの向上のために、エージェントはそのようなインタラクションを介して、ユーザやエージェント社会全体にとっての適切な行動を学習する必要がある。

しかし、ユーザや個人の代理であるエージェントが自身の利益を優先して行動すると、社会全体の利益を下げるジレンマ的な状況が存在する。このようなジレンマをモデル化したゲームが囚人のジレンマゲームである。囚人のジレンマゲームでは、全員が協調行動を取ることがパレート最適であり、社会全体の利益を高める。しかし、個人の利益を優先すると裏切り行動を取ることになり、これがナッシュ均衡となる。囚人のジレンマでは、エージェントが社会全体として協調する必要がある。

囚人のジレンマゲームにおいて、協調を促進させる学習手法についての研究 [1], [2] がある。[1] は、エージェントの得る報酬を主観的報酬とした Q 学習手法を提案した。客観的報酬から生成される主観的報酬では協調が支配戦略となるため、相互協調が発生する。しかし、隣接エージェントそれぞれについて Q 値を持たず、全体で 1 つの Q 値としているため、社会的な振る舞いについて考慮していない。[2] は、強化学習戦略を提案し、特定の条件下で相互協調が発生することを示した。過去の利得を高く評価する場合に、強化学習戦略はしつぱ返し戦略のように振る舞い、相互協調が発生する。しかし、この研究ではエージェント数は 2 体であり、エージェントが社会を構成しているとはいえない。

著者らは、囚人のジレンマゲームにおいて、ネットワーク全体に協調を促進させる研究 [3], [4], [5] を行っている。[4] では、協調期待戦略を提案し、すべてのエージェントが協調期待戦略を行う場合に、ネットワーク全体に協調が促進することを示した。協調期待戦略を取るエージェントに、自身の利益のみを考え合理的に行動するエージェントを混ぜた研究 [5] がある。合理的エージェントは行動の結果を学習することで、行動がナッシュ均衡である裏切りに収束する。その結果、ネットワーク全体に裏切りが広がり、協調期待戦略を取るエージェントは協調行動を維持できな

a) k.muranaka@isl.cs.waseda.ac.jp

b) t.otsuka@isl.cs.waseda.ac.jp

c) sugawara@waseda.jp

くなることを示した。これに対し、協調期待戦略を拡張し、裏切り行動を取るエージェントに対する頑健性を高めた研究 [3] がある。この研究では、強化協調期待戦略を提案し、裏切りエージェントの混在する状況においても、協調行動を維持できることを示した。これらの研究では、エージェントの関係をネットワークで、インタラクションをゲームの利得でモデル化している。

また、囚人のジレンマゲームにおいて、異質なエージェントを混在させたときの影響を調査した研究がある。[6] は、エージェントに年齢を導入し、年齢によって利得構造を変化させることで、異質なエージェントが混在するネットワークを実現した。この研究では、利得構造を動的に変化させることにより、特定の状況下では協調の進化が促進されることを示した。また、利得構造の違いに着目し、その影響を調査した研究 [7], [8] もある。[7] では、2人囚人のジレンマゲームにおいて、対称ゲームと非対称ゲームの違いを調査した。インタラクションを行う者同士の利得構造が対称ゲームでは同じであるが、非対称ゲームでは異なる。利得構造が非対称の場合には協調行動が選択されづらくなることを示した。また、[8] では、非対称の利得構造において、同時ゲームと順次ゲームでの協調選択率を調査した。同時ゲームにおいては、利得構造の非対称性は協調選択率を低下させ、順次ゲームにおいては、行動を決定する順番が協調選択率に影響することを示した。

本研究では、異なる利得行列を持つエージェントが混在するネットワークにおける強化協調期待戦略の頑健性を調査する。[3], [4], [5] では、すべてのエージェントが囚人のジレンマゲームの利得行列を持つとしていたが、実社会ではエージェントの行動に関する価値観が同一であるとは限らない。そのため、[3] では、常に協調しないあるいは確率的に協調しないエージェントを混在させ、協調促進の頑健性を示していた。また、[6], [7], [8] では、利得構造を変化させることでエージェントの価値観を変化させている。そこで、本研究では、エージェントの価値観を変化させる手法として、利得構造を変化させることに着目し、一部のエージェントの利得構造を反調整ゲームとする。反調整ゲームを行うエージェントを囚人のジレンマゲームを行うエージェントに混在させ、その混在率やネットワーク構造が強化協調期待戦略の頑健性に与える影響を調査する。

## 2. 準備

### 2.1 ネットワークモデル

本節では、本研究で用いるネットワーク構造について説明する。本研究では、完全グラフ、正則グラフ、ランダムグラフ [9] の3つのネットワークを用いる。

#### 2.1.1 完全グラフ

完全グラフは、全てのノード間にリンクが存在するグラフである。

#### 2.1.2 正則グラフ

正則グラフは、各ノードの次数が全て同じであるグラフである。本研究では、ノードを円状に並べ、隣接する  $m$  個のノードに対してリンクを生成する。すなわち、ネットワークの平均次数  $k$  は  $m$  となる。例えば  $m = 10$  とすると、ノードの両隣に5つずつリンクを生成し、 $k = 10$  となる。

#### 2.1.3 ランダムグラフ

ランダムグラフは、完全グラフや正則グラフとは異なり、生成する度にネットワーク構造が異なるモデルである。このネットワークには、リンクを張る確率  $p(0 < p < 1)$  というパラメータがあり、各ノード間に  $p$  の確率でリンクを張ることで生成される。[9]

## 2.2 ゲーム理論

### 2.2.1 囚人のジレンマゲーム

囚人のジレンマゲームは2人ゲームであり、協調と裏切りの2種類の行動を取ることができる。ここで、2種類の行動を  $S = C$  (協調),  $D$  (裏切り) と定義する。囚人のジレンマの利得行列は以下のように定義される。

	$C$	$D$
$C$	$R$	$S$
$D$	$T$	$P$

また、利得行列の各値  $R, S, T, P$  について、次の関係が成り立つ。

$$T > R > P > S \text{ and } 2R > T + S.$$

パレート最適な行動の組み合わせは  $(C, C)$  であり、全体の利得が最も大きい。しかし、支配戦略として  $D$  が存在し、 $(D, D)$  がナッシュ均衡となる。従って、両者が合理的な戦略を取るとナッシュ均衡に陥る。

### 2.2.2 反調整ゲーム

反調整ゲームは2人ゲームであり、2種類の行動を取ることができる。ここで、2種類の行動を  $S = C, D$  と定義する。これに対して、利得行列は以下のように定義される。

	$C$	$D$
$C$	$R$	$S$
$D$	$T$	$P$

また、利得行列の各値  $R, S, T, P$  について、次の関係が成り立つ。

$$R < T \text{ and } S > P.$$

パレート最適な行動の組み合わせは  $(C, D), (D, C)$  の2つであり、全体の利得が最も大きい。また、ナッシュ均衡も  $(C, D), (D, C)$  の2つである。

## 3. 行動戦略

本章では、エージェントが採用する行動戦略について説

明する。エージェントはインタラクションの際に自身の行動を決定する必要がある。このとき、エージェントは行動戦略に従って行動を決定する。

### 3.1 統合法

統合法 [4] は、合理的に行動し自身の利益を最大化しようとする戦略である。この戦略は囚人のジレンマゲームと反調整ゲームの両方に適用される戦略である。エージェント  $i$  は強化学習を行い、その結果から各隣接エージェント  $j$  に対する自身の行動を決定する。エージェント  $i$  は各隣接エージェント  $j$  と行動  $s \in S$  に対して  $Q$  値を保持している。エージェント  $i$  のラウンド数  $t$  における  $Q_t^i(j, s)$  は、学習率を  $\alpha$ 、前ラウンドでのエージェント  $j$  とのインタラクションの報酬を  $r_{t-1}^i(j)$  として次の式で表される。

$$Q_t^i(j, s) = (1 - \alpha)Q_{t-1}^i(j, s) + \alpha r_{t-1}^i(j) \quad (1)$$

また、ラウンド  $t$  におけるエージェント  $i$  の基本行動  $a_i(t)$  は次の式で決定される。自身の利益を最大化させる行動が基本行動である。

$$a_i(t) = \arg \max_{s \in S} p_i(s) \quad (2)$$

ここで、エージェント  $i$  の隣接エージェントの集合を  $N_i$  として、優先度  $p_i(s)$  は次の式で定義される。

$$p_i(s) = \sum_{j \in N_i} \delta(s, a_j^i(t-1)) \quad (3)$$

また、 $\delta$  は  $s_1, s_2 \in S$  について次の値を取る関数である。

$$\delta(s_1, s_2) = \begin{cases} 1 & \text{if } (s_1 = s_2) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

なお、行動ごとの優先度が同値の場合には、基本行動  $a_i$  をランダムに選択する。学習戦略には  $\epsilon$ -greedy 法を採用し、インタラクションごとに判定を行い、実際の行動として基本行動またはランダムに行動を選択する。

### 3.2 協調期待戦略

協調期待戦略 [4] (Expectation of Cooperation 戦略, 以降, EoC 戦略) では、自身が協調試行期間にある場合には必ず協調行動を取る戦略である。この戦略を採用するエージェントは、 $(C, C)$  がパレート最適であると認識しており、周囲に協調行動が広がることを期待する。本研究で用いるゲームの内、 $(C, C)$  がパレート最適となるのは囚人のジレンマゲームのみである。そのため、EoC 戦略は囚人のジレンマゲームの利得行列を持つエージェントのみに対して適用される。エージェントは協調試行期間を表す  $l$  という値を保持しており、 $l > 0$  のときに協調試行期間にあるとす

る。インタラクションの際に、相互協調が発生するとエージェントは  $l = L$  と設定する。ここで、 $L$  は正の整数である。すなわち、相互協調が発生すると協調試行期間に入り、その後  $L$  回の間で協調行動をとるということである。協調試行期間中のインタラクションにおいて、相互協調が発生した場合には  $l = L$  と再設定するが、相互協調が発生しなかった場合には  $l$  を 1 減少させる。なお、エージェントが協調試行期間にない場合には、3.1 節の統合法に基づき行動を決定する。また、 $l$  の初期値は 0 である。

### 3.3 強化協調期待戦略

強化協調期待戦略 [3] (Enhanced Expectation of Cooperation 戦略, 以降, EEoC 戦略) は、3.2 節の EoC 戦略を拡張した戦略である。EoC 戦略と同様に、EEoC 戦略は囚人のジレンマゲームの利得行列を持つエージェントのみに対して適用される。エージェントは各隣接エージェント  $j$  に対して  $d_j$  という値を保持している。これは、各隣接エージェント  $j$  が直前のインタラクションまでで連続して裏切りを選択した回数を表す値である。エージェントは自身が協調試行期間にある間、各隣接エージェント  $j$  の行動を記憶し、インタラクション毎に  $d_j$  を更新する。また、裏切り判定回数  $M$  という値を持ち、 $d_j \geq M$  である隣接エージェント  $j$  を裏切りエージェントと判別する。ここで、 $M$  は正の整数である。裏切りエージェントとのインタラクションでは、協調ではなく統合法に基づく基本行動を選択する。なお、隣接エージェント  $j$  が協調を選択した場合には、 $d_j = 0$  と設定する。また、 $d_j$  の初期値は 0 である。

## 4. 実験と考察

本章では、本研究で行う実験の内容とその結果について述べる。

### 4.1 実験内容

ネットワーク上にエージェントを配置し、エージェントの行動がインタラクションを通してどのように推移するのかを調査する。3.1 節の統合法に基づいて自身の行動を決定するエージェントを合理的エージェントとし、3.3 節の EEoC 戦略を取るエージェントを EEoC エージェントとする。本研究では、価値観の異なる合理的エージェントに対する EEoC 戦略の頑健性を調査するために、EEoC エージェントに合理的エージェントを混ぜた場合における、EEoC エージェントの行動の推移を調査する。このとき、ネットワーク全体の価値観が均一の場合と均一でない場合を比較するため、合理的エージェントがそれぞれ囚人のジレンマゲームと反調整ゲームの利得行列を持つ場合について実験する。また、EEoC エージェントと合理的エージェントの割合はあらかじめ固定し、シミュレーションごとに 10% ずつ割合を変更させる。

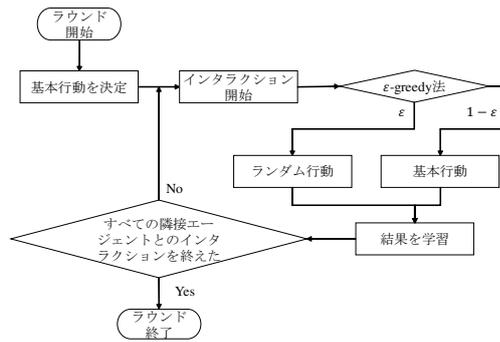


図1 シミュレーションの流れ

ネットワーク上にエージェントを配置するにあたって、エージェントはノードと1対1に対応させる。また、リンクがエージェント同士の関係性を表し、リンクのあるエージェント(隣接エージェント)とインタラクションを行う。インタラクションはゲーム理論(2.2節)でモデル化し、自身の行動と相手の行動に対応する利得を報酬として貰う。

#### 4.1.1 シミュレーションの流れ

シミュレーションの流れについて説明する。すべてのエージェントがすべての隣接エージェントとのインタラクションを1度ずつ行うまでを1ラウンドとする。このラウンドを繰り返すことでシミュレーションを行う。なお、エージェントはシミュレーション全体を通して、自身の戦略を途中で変更することはない。自身の行っている利得行列も変わることはない。1ラウンドのシミュレーションの流れを1体の合理的エージェントに着目して説明する。1ラウンドのシミュレーションの流れを図1に示す。エージェントはラウンド開始時に統合法により自身の基本行動を決定する。その後、隣接エージェントとインタラクションを開始する。このとき、どの隣接エージェントとインタラクションを行うかはランダムに決定される。インタラクションの際に選択する行動は、 $\epsilon$ -greedy法により決定され、基本行動かランダム行動を行う。そして、インタラクションを終えて報酬を貰い、報酬を基に結果を学習する。インタラクションを繰り返し、すべての隣接エージェントとのインタラクションを終えるとラウンドを終了する。

#### 4.2 実験結果と考察

2.1節で述べた各ネットワークについて、EEoCエージェントに囚人のジレンマゲームの利得行列を持つ合理的エージェントを混ぜた場合と、反調整ゲームの利得行列を持つ合理的エージェントを混ぜた場合の結果を示す。以下の結果は独立試行20回の平均である。なお、どちらの場合においても、EEoCエージェントは囚人のジレンマゲームの利得行列を持つ。実験に用いたパラメータを表1に示す。囚人のジレンマゲームの利得構造の値は、 $R = 3, S = 0, T = 5, P = 1$ とし、反調整ゲームの利得構造の値は、 $R = -1, S = 1, T = 1, P = -1$ とした。また、

強化学習に用いる  $Q$  値の初期値は6とした。本研究では、[4]に習い、相互協調率が0.9以上の場合に協調の促進が成功したと判断する。

表1 パラメータ

パラメータ	値
エージェント数(完全グラフ)	300
エージェント数(その他)	1000
ラウンド数	5000
平均次数 $k$ (正則グラフ)	10
リンクを張る確率 $p$ (ランダムグラフ)	0.01001
学習率 $\alpha$	0.1
$\epsilon$ ( $\epsilon$ -greedy法)	0.05
協調試行期間 $L$	3
裏切り判定回数 $M$	2

#### 4.2.1 完全グラフ

ネットワーク構造が完全グラフの場合における実験結果を示す。合理的エージェントの利得行列を囚人のジレンマゲームの利得行列にした場合の、EEoCエージェントに混ぜる合理的エージェントの割合を10%から40%に変化させたときの、EEoCエージェント間の相互協調率を図2に示す。なお、図には掲載していないが、合理的エージェントの割合が0%の場合は10%の場合と、50%以上の場合は40%と同様の結果であった。また、合理的エージェントの利得行列を反調整ゲームの利得行列にした場合の、EEoCエージェントに混ぜる合理的エージェントの割合を10%から50%と、80%と90%に変化させたときの、EEoCエージェント間の相互協調率を図3に示す。なお、図には掲載していないが、合理的エージェントの割合が0%の場合は10%の場合と、60%と70%の場合は50%と同様の結果であった。

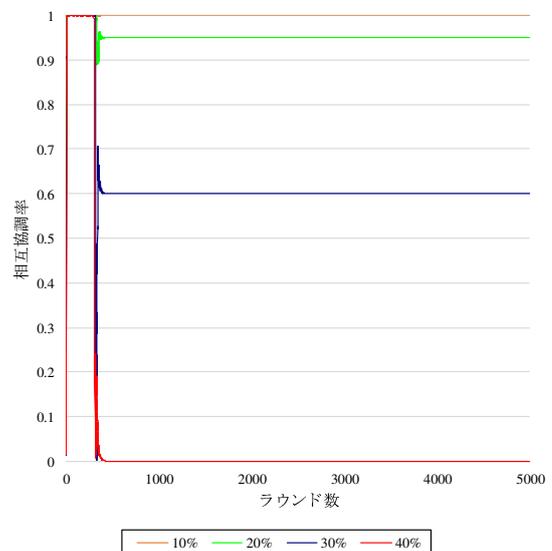


図2 完全グラフにおける相互協調率(囚人のジレンマゲーム)

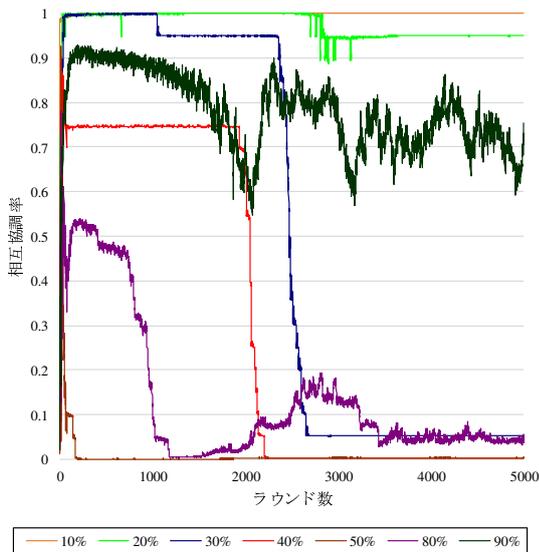


図 3 完全グラフにおける相互協調率 (反調整ゲーム)

図 2 と図 3 を見ると、合理的エージェントの割合が 20% 以下の場合には、合理的エージェントのゲーム構造に関係なく、相互協調率が 0.9 以上に収束し、協調の促進に成功している。しかし、合理的エージェントの割合が 30% 以上の場合には、合理的エージェントのゲーム構造に関係なく、相互協調率は 0.9 を下回り、協調の促進に失敗している。また、合理的エージェントの割合が 30% の場合には、合理的エージェントのゲーム構造による違いが現れている。合理的エージェントが囚人のジレンマゲームの利得行列を持つ場合では、相互協調率は 0.6 付近に収束している。

一方、合理的エージェントが反調整ゲームの利得行列を持つ場合では、相互協調率は初めは 0.9 以上を維持しているが、ラウンド 2400 付近を境に急激に 0 付近に収束する。同様に、反調整ゲームの利得行列を持つ合理的エージェントの割合が 40% の場合においても、相互協調率は初めは 0.7 付近の値を維持しているが、ラウンド 2000 付近を境に 0 付近に収束する。

この結果は、ゲーム構造によってナッシュ均衡が異なることが原因であると考えられる。囚人のジレンマゲームにはナッシュ均衡が 1 つしか存在せず、合理的エージェントの行動は裏切りに収束する。一方、反調整ゲームにはナッシュ均衡が 2 つ存在し、合理的エージェントの行動は 1 つに定まらない。つまり、反調整ゲームの利得行列を持つ合理的エージェントは自身の行動を変更することがある。このため、EEoC エージェントは合理的エージェントを裏切りエージェントと断定することができなくなる。その結果、協調試行期間を維持できなくなり、自身も裏切り行動に収束してしまうと考えられる。

しかし、反調整ゲームの利得行列を持つ合理的エージェントの割合がさらに増大すると、相互協調率は 0 付近に収束せず増大する。反調整ゲームの利得行列を持つ合理的

エージェントの割合が 80% の場合には、ラウンド 2800 付近で、相互協調率が 0.2 付近まで増大している。また、反調整ゲームの利得行列を持つ合理的エージェントの割合が 90% の場合には、相互協調率はラウンド全体を通して 0.6 から 0.9 程度の値を維持している。これは、反調整ゲームの利得行列を持つ合理的エージェントの行動が、エージェントごとに協調または裏切りに収束していることが原因であると考えられる。合理的エージェントの行動が住み分けにより固定されることにより、EEoC エージェントは裏切りエージェントの判別をしやすくなる。この結果、相互協調率が増大すると考えられる。

以上より、合理的エージェントのゲーム構造の違いに関係なく、合理的エージェントの割合が 20% 以下の場合に協調の促進に成功し、30% 以上の場合には協調の促進に失敗することが分かった。つまり、ネットワーク全体の価値観が均一でない状況での EEoC 戦略の頑健性は、ネットワーク全体の価値観が均一の場合と同程度の頑健性であると考えられる。

#### 4.2.2 正則グラフ

ネットワーク構造が正則グラフの場合における実験結果を示す。合理的エージェントの利得行列を囚人のジレンマゲームの利得行列にした場合の、EEoC エージェントに混ぜる合理的エージェントの割合を 10% から 70% に変化させたときの、EEoC エージェント間の相互協調率を図 4 に示す。なお、図には掲載していないが、合理的エージェントの割合が 0% の場合は 10% の場合と、80% 以上の場合には 70% と同様の結果であった。また、合理的エージェントの利得行列を反調整ゲームの利得行列にした場合の、EEoC エージェントに混ぜる合理的エージェントの割合を 10% から 70% に変化させたときの、EEoC エージェント間の相互協調率を図 5 に示す。なお、図には掲載していないが、合理的エージェントの割合が 0% の場合は 10% の場合と、80% 以上の場合には 70% と同様の結果であった。

図 4 と図 5 を見ると、合理的エージェントの割合が 20% 以下の場合には、合理的エージェントのゲーム構造に関係なく、相互協調率が 0.9 以上に収束し、協調の促進に成功している。しかし、合理的エージェントの割合が 30% 以上の場合には、合理的エージェントのゲーム構造に関係なく、相互協調率は 0.9 以上に収束せず、協調の促進に失敗している。特に、合理的エージェントの割合が 60% 以上の場合には、0 付近に収束している。また、全体として合理的エージェントのゲーム構造が反調整ゲームである場合の方が、相互協調率がわずかに大きい。

この結果は、ゲーム構造によってナッシュ均衡が異なることが原因であると考えられる。囚人のジレンマゲームにはナッシュ均衡が 1 つしか存在せず、合理的エージェントの行動は裏切りに収束する。一方、反調整ゲームにはナッシュ均衡が 2 つ存在し、合理的エージェントの行動は 1 つ

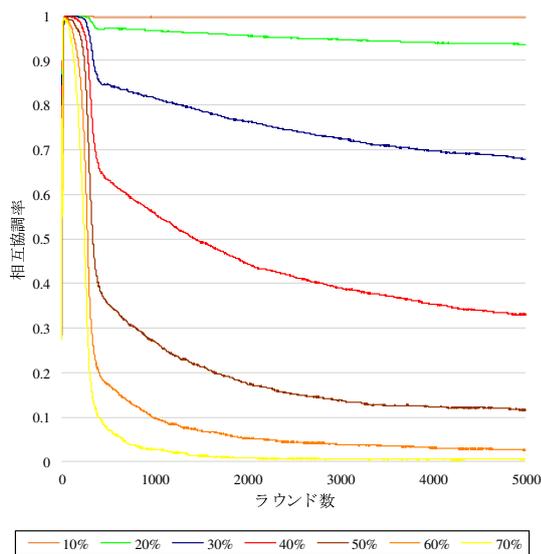


図 4 正則グラフにおける相互協調率 (囚人のジレンマゲーム)

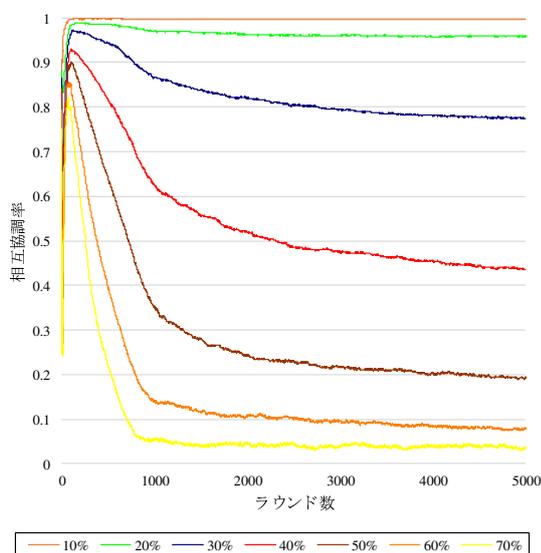


図 5 正則グラフにおける相互協調率 (反調整ゲーム)

に定まらない。つまり、反調整ゲームの利得行列を持つ合理的エージェントは裏切りではなく協調行動を取ることがある。このため、EEoC エージェント以外にも協調行動を取るエージェントが存在することになる。つまり、協調が促進されることになり、相互協調率が大きくなると考えられる。

以上より、合理的エージェントのゲーム構造の違いに関係なく、合理的エージェントの割合が20%以下の場合に協調の促進に成功し、30%以上の場合には協調の促進に失敗することが分かった。つまり、ネットワーク全体の価値観が均一でない状況でのEEoC戦略の頑健性は、ネットワーク全体の価値観が均一の場合と同程度の頑健性であると考えられる。

#### 4.2.3 ランダムグラフ

ネットワーク構造がランダムグラフの場合における実験

結果を示す。合理的エージェントの利得行列を囚人のジレンマゲームの利得行列にした場合の、EEoC エージェントに混ぜる合理的エージェントの割合を10%から50%に変化させたときの、EEoC エージェント間の相互協調率を図6に示す。なお、図には掲載していないが、合理的エージェントの割合が0%の場合は10%の場合と、60%以上の場合は50%と同様の結果であった。また、合理的エージェントの利得行列を反調整ゲームの利得行列にした場合の、EEoC エージェントに混ぜる合理的エージェントの割合を10%から50%に変化させたときの、EEoC エージェント間の相互協調率を図7に示す。なお、図には掲載していないが、合理的エージェントの割合が0%の場合は10%の場合と、60%以上の場合は50%と同様の結果であった。

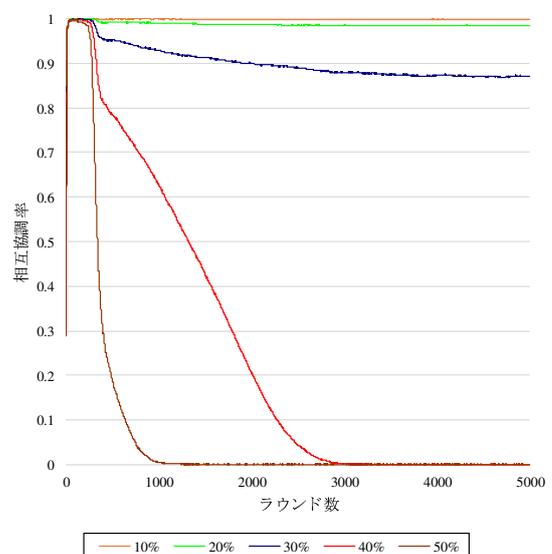


図 6 ランダムグラフにおける相互協調率 (囚人のジレンマゲーム)

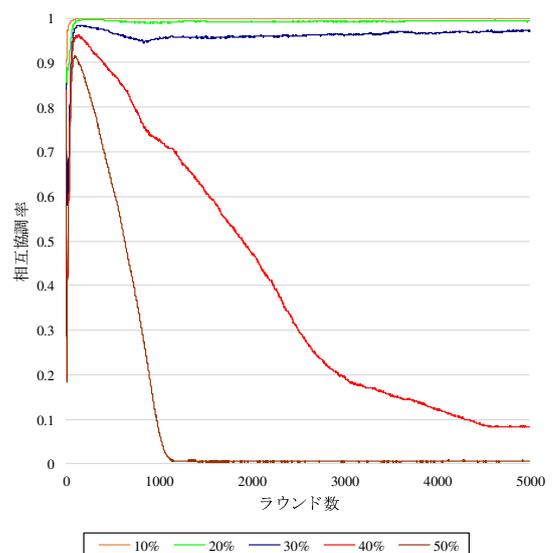


図 7 ランダムグラフにおける相互協調率 (反調整ゲーム)

図6と図7を見ると、合理的エージェントの割合が20%

以下の場合には、合理的エージェントのゲーム構造に関係なく、相互協調率が0.9以上に収束し、協調の促進に成功している。しかし、合理的エージェントの割合が40%以上の場合には、合理的エージェントのゲーム構造に関係なく、相互協調率は0.9以上に収束せず、協調の促進に失敗している。また、合理的エージェントの割合が30%の場合には、合理的エージェントのゲーム構造が反調整ゲームである場合には、相互協調率が0.9以上に収束しており、協調の促進に成功している。一方、合理的エージェントのゲーム構造が囚人のジレンマゲームである場合には、相互協調率が0.9未満に収束しており、協調の促進に失敗している。また、全体として合理的エージェントのゲーム構造が反調整ゲームである場合の方が、相互協調率がわずかに大きい。

この結果は、ゲーム構造によってナッシュ均衡が異なることが原因であると考えられる。囚人のジレンマゲームにはナッシュ均衡が1つしか存在せず、合理的エージェントの行動は裏切りに収束する。一方、反調整ゲームにはナッシュ均衡が2つ存在し、合理的エージェントの行動は1つに定まらない。つまり、反調整ゲームの利得行列を持つ合理的エージェントは裏切りではなく協調行動を取ることがある。このため、EEoCエージェント以外にも協調行動を取るエージェントが存在することになる。つまり、協調が促進されることになり、相互協調率が大きくなると考えられる。

以上より、合理的エージェントのゲーム構造の違いに関係なく、合理的エージェントの割合が20%以下の場合に協調の促進に成功し、40%以上の場合には協調の促進に失敗することが分かった。また、合理的エージェントの割合が30%の場合には、合理的エージェントが反調整ゲームの利得行列を持つ場合には協調の促進に成功し、囚人のジレンマゲームの利得行列を持つ場合には協調の促進に失敗している。つまり、EEoC戦略はネットワーク全体の価値観が均一の状況よりも、ネットワーク全体の価値観が均一でない状況での頑健性をより有していると考えられる。

## 5. 結論

本研究では、反調整ゲームの利得行列を持つ合理的エージェントが混在するネットワークにおけるEEoC戦略の協調促進の頑健性について調査した。4.2節の考察から、ネットワーク構造が完全グラフ、正則グラフの場合には、ネットワーク全体の価値観が均一の場合と均一でない場合とで、協調の促進に成功する合理的エージェントの割合は同じであった。すなわち、EEoC戦略の協調促進の頑健性は、ネットワーク全体の価値観が均一の場合と均一でない場合とで同程度であると考えられる。一方、ランダムグラフにおいては、ネットワーク全体の価値観が均一の場合と均一でない場合とで、協調の促進に成功する合理的エージェントの割合は異なった。ネットワーク全体の価値観が均一で

ないときの方が、より多くの合理的エージェントを含む場合においても協調の促進に成功することがあった。すなわち、ネットワーク全体の価値観が均一である場合よりも、ネットワーク全体の価値観が均一でない場合の方が、EEoC戦略は協調促進の頑健性を有していると考えられる。

今後の課題としては、ゲーム構造を変更することが挙げられる。本研究では、反調整ゲームの利得行列を持つエージェントに対する頑健性を評価したが、チキンゲームやタカハトゲームの利得行列を持つエージェントに対する頑健性を評価することが考えられる。また、ネットワーク構造に現実世界のネットワークを適応させることも挙げられる。つまり、現実世界に近い環境でのEEoC戦略の協調促進の頑健性を調査することができる。そのためには、FacebookやTwitterなどのソーシャルネットワーキングサービスが成すネットワークにて実験を行うことが考えられる。

## 参考文献

- [1] 森山甲一, 栗原聡, and 沼尾正行. 囚人のジレンマゲームにおける主観的効用の進化にパラメータが与える影響について. In *Proceedings of Joint Agent Workshops and Symposium*, 2011.
- [2] 鳥居拓馬 and 日高昇平. 利己的な相互協調: 強化学習主体による不確実な囚人のジレンマ. 人工知能学会全国大会, 2015. 1L5-3.
- [3] T. Otsuka and T. Sugawara. Promotion of robust cooperation among agents in complex networks by enhanced expectation-of-cooperation strategy. In *Proceedings of Complex Networks*, volume 689, pages 815–828. Springer, 2017.
- [4] R. Shibusawa, T. Otsuka, and T. Sugawara. Spread of cooperation in complex agent networks based on expectation of cooperation. In *Proceedings of Principles and Practice of Multi-Agent Systems*, volume 9862, pages 76–91. Springer, 2016.
- [5] 大塚知亮 and 菅原俊治. 協調期待戦略による協調促進の頑健性について. In *Proceedings of Joint Agent Workshops and Symposium*, pages 87–92, 2016.
- [6] Golriz Rezaei and Michael Kirley. Heterogeneous payoffs and social diversity in the spatial prisoner's dilemma game. In *Proceedings of Simulated Evolution and Learning*, pages 585–594. Springer, 2008.
- [7] J.P. Sheposh and JR P.S. Gallo. Asymmetry of payoff structure and cooperative behavior in the prisoner's dilemma game. *Journal of Conflict Resolution*, 17:321–333, 1973.
- [8] T. K. Ahn, Myungsuk Lee, Lore Ruttan, and James Walker. Asymmetric payoffs in simultaneous and sequential prisoner's dilemma games. *Public Choice*, 132:353–366, 2007.
- [9] P. Erdos and A. Renyi. On random graphs, i. *Publicationes Mathematicae*, 6:290–297, 1959.