



必ずしも十分であるとはいえない．Rapoport et al. [27] は 2 種類の実験を行い，金銭報酬の大きさが被験者の行動に影響することを明らかにした．また，McKelvey and Palfrey [20] は，被験者が適応的に戦略選択していると考え，確率的行動モデルである AQRE による説明を試みた．Kawagoe and Takizawa [15] は，被験者の戦略選択における合理性の度合いをパラメータ  $k$  により表現する Level- $k$  理論に基づき，被験者の行動の説明を試み，Agent Quantal Response Equilibrium (AQRE) モデルに比べて被験者の平均的な行動を説明することができることを示した．ムカデゲーム以外のゲームの被験者実験に対しても，均衡概念だけでは説明できない結果がいくつか報告されているが [4, 5, 8, 19, 21, 22, 26, 27, 30, 34]，これらに対して被験者の意思決定におけるノイズあるいはエラーによる説明が試みられている [8, 19, 26]．Ochs [26] は，一般化コイン一致ゲームにおける被験者の行動に対して，ランダム性を導入したモデルによる説明を試みた．この結果に対して McKelvey and Palfrey [19] は，高い期待利得をもつ行動は選択されやすいと解釈し，意思決定にノイズあるいはエラーを含むことを考慮した均衡概念を提案している [8]．このように，多くの被験者実験における被験者の行動を説明するためのさまざまなモデルが提案されている．

ゲームの実験における被験者の決定には，様々な要因が影響していると考えられるが，これらの要素は被験者の行動に複合的に影響していると考えられる．本論文ではムカデゲームにおける被験者の行動を説明する上で，さらにこれらの各要素の影響を分析する．一般に，均衡理論ではプレイヤーは利己的で合理的な意思決定をすると仮定されている．しかし，この仮定が現実の人間の意思決定と必ずしも整合しているとはいえないことが部分ゲーム完全均衡と被験者の行動が一致しない理由として考えられる．さらに，実験が繰返し行われていることに着目し，人間は試行錯誤的で相手の行動や自身の経験を参考にしながら長期的な視野に基づいた意思決定を行っていると考え，様々な行動要素を導入したモデルを形成しやすいニューラルネットワークと遺伝的アルゴリズムに基づいた意思決定と学習機構を持つ人工適応型エージェントを用いたシミュレーション分析を行う．

本論文の構成は以下の通りである．2 節ではムカデゲームについて，3 節では人工適応型エージェントを用いたシミュレーションモデルについて，4 節ではシミュレーションの結果について説明を行い，5 節で本論文の成果をまとめる．

## 2. ムカデゲーム

部分ゲーム完全均衡は，展開型ゲームに対する均衡解概念として知られている．均衡概念ではプレイヤーの利得を最大化するという意味での合理的な意思決定が仮定されており，複数のプレイヤーがあらかじめ決められた順番で意思決定する展開型ゲームでは，後ろ向き帰納法により部分ゲーム完全均衡を導出することができる．展開型ゲームの利得構造によっては，各プレイヤーが個々の意思決定ノードにおいて最適応答戦略を選択した場合であっても，パレート最適性を満たさない利得を得ることがある．このような場合には，プレイヤーが協調して特定の戦略を選択することで，より大きな利得を得ることも可能である．Rosenthal [28] は，このような性質を持つ展開型ゲームとして，ムカデゲームを考案した．ムカデゲームでは，プレイヤーが事前に決められた順番で *pass* か *take* かを選択する．*pass* を選択すると次のプレイヤーの意思決定の手番となる．あるノードでプレイヤーが *take* を選択すると，ゲームが終了し，各プレイヤーは対応する終端ノードに割り当てられた利得を受け取る．3 人ムカデゲームの例を図 1 に示す．ムカデゲームにおいて，プレイヤーが *take* を選択することで得られる利得の 3 項組が図 1 の各ノードの下の終端に示されており，3 項組の上からプレ

ィヤー 1, 2, 3 の利得である．各ノードの上には意思決定を行うプレイヤーが示されており， $p_1, p_2, p_3$  はそれぞれプレイヤー 1, 2, 3 を示す．ノードの中の数はノードの番号である．

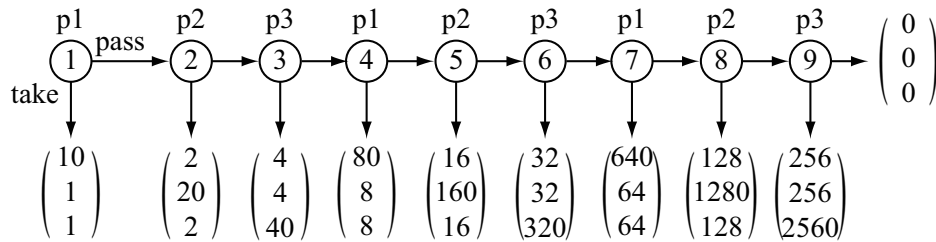


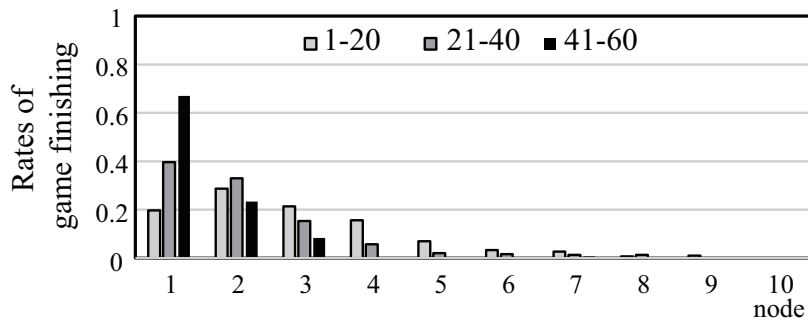
図 1: Game tree: a centipede game with three players

展開型ゲームにおいて，あるノードから先を独立した1つの展開型ゲームとみなしたゲームをサブゲームといい，すべてのサブゲームで各プレイヤーが最適応答戦略を選択するような戦略集合を部分ゲーム完全均衡という．図1のムカデゲームでは，次に示す部分ゲーム完全均衡は後ろ向き帰納法により導出することができる．図1のゲームでは，最終ノードであるプレイヤー3の手番のノード9では *take* (利得 2560) が最適応答となるので，プレイヤー3は *take* を選択してゲームを終了する．次に，プレイヤー3は *take* を選択すると仮定して，プレイヤー2は *take* か *pass* を選択する．ノード8でのプレイヤー2も同様に *take* (利得 1280) が最適応答となるので，プレイヤー2は *take* を選択してゲームを終了する．同様の手順をすべてのノードで行うと，プレイヤー1がノード1で *take* を選択して，各プレイヤーが利得 (10, 1, 1) を得ることが部分ゲーム完全均衡による予測となる．このことから，各プレイヤーが利得を最大化する合理的な選択をすると仮定しているにも関わらず，部分ゲーム完全均衡であるノード1で *take* を選択してゲームが終了する場合に比べて，より大きな番号のノードでプレイヤーが *take* を選択した場合の利得のほうが大きい．この意味で，部分ゲーム完全均衡の実現は協調に失敗していると解釈することができ，大きい番号のノードでゲームを終了することが，協調に成功しているといえる．ここで，ノード1, 2, 3でゲームが終了することは，1度も *pass* を選択しないプレイヤーが存在することからプレイヤーは非協調関係にあると解釈することができ，ノード4以降でゲームが終了することはすべてのプレイヤーが一度は *pass* を選択するので協調関係にあるといえる．これまでの被験者実験の結果 [18, 22, 27] において，被験者の行動は部分ゲーム完全均衡の予測から逸脱しており，協調関係が観測されている．次の小節で Rapoport et al. [27] による3人ムカデゲームの被験者実験について述べる．

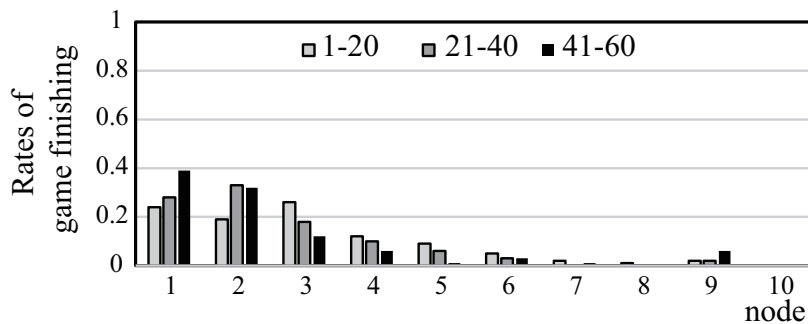
### 2.1. Rapoport et al. の被験者実験 [27]

Rapoport et al. は，図1に示される3人ムカデゲームの被験者実験を行った．被験者は15人ごとの4つのグループに分けられ，それぞれ60回のゲームが行われた．各グループでの60回のゲームをセッションとよぶ．各回のゲームでは15人をランダムに3人ずつの5つのゲームグループに分ける．各ゲームグループでは各被験者にランダムにそれぞれ1, 2, 3の番号を割り当て，これをプレイヤー番号とした．実験結果として，3つのセッションではノード1, 2, 3で *take* が選択されゲームが終了する割合が高く，他の1つのセッションではノード1, 2, 3で *take* が選択される割合が低かった．本論文では，この被験者実験を実験 E-I とし，ノード1, 2, 3でゲームが終了する割合が高い3つのセッションでの結果を実験結果 E-I-A，残り1つのセッションでの結果を実験結果 E-I-B とする．実験結果 E-I-A, E-I-B において，

ノード1で *take* が選択され、部分ゲーム完全均衡からの予測と整合的となる割合の平均はそれぞれ 42.2% と 30.3% であった。実験結果 E-I-A, E-I-B において横軸をノード、縦軸を 1-20 期, 21-40 期, 41-60 期でゲームが終了した割合を図2に示す。



(a) Experimental Result E-I-A



(b) Experimental Result E-I-B

図2: Experimental result E-I-A and E-I-B [27]

図2より、実験の41期以降でノード1でゲームが終了した割合は、実験結果 E-I-A では 0.667 であるのに対して、実験結果 E-I-B では 0.390 であった。また、最終のノード9で終了したゲームの割合は、実験結果 E-I-A では各セッションで 0.017 であったのに対して、実験結果 E-I-B では 0.167 であった。特に実験結果 E-I-B において、ある特定の5人の被験者が同じグループでゲームを行ったとき、ノード9でゲームが終了する割合が高かった。このように、被験者の中に協調傾向が強く、積極的に *pass* を選択する被験者が存在することが確認できる。実験結果 E-I-B ではこれらの被験者の影響により、ゲームを繰り返し行っても部分ゲーム完全均衡からの予測と整合的となる頻度が低いままであったと考えられる。

さらに、Rapoport et al. は被験者に支払われる金銭報酬の大きさが実験の結果に影響すると考え、金銭報酬を実験 E-I の 1/100 とした実験を行った。グループ数や実験期間など金銭以外は実験 E-I と同様であった。この被験者実験を実験 E-II、結果を実験結果 E-II とする。実験結果 E-II として、ノード 1, 2, ..., 9 でのゲーム終了割合を図3に示す。

図3より、実験 E-I に比べてノード 1-3 での終了割合が小さく、被験者間に協調関係が構築されていたと解釈できる。

Rapoport et al. [27] は以上2種類の実験を行い、金銭報酬が大きければサブゲーム完全均衡と整合した行動が被験者により選択されやすくなり、逆に小さければ協調行動が選択されやすくなる傾向があることを示し、金銭報酬の大きさが被験者の行動に影響を与えることを明らかにした。

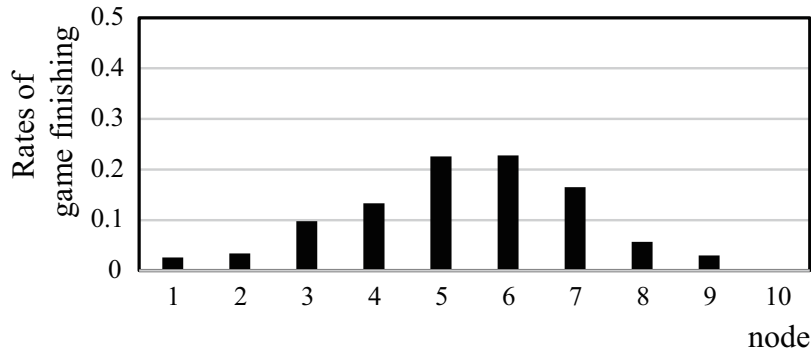


図 3: Rate of the games finishing (Experiment E-II [27])

## 2.2. ムカデゲームの確率的行動選択モデル

被験者はゲームの利得に基づいて確率的に戦略を選択していると考え、ムカデゲームにおける被験者行動の説明のための行動選択モデルが提案されている。

McKelvey and Palfrey [20] は、Harsanyi [9] の確率的選択モデルを修正し、プレイヤー  $i$  が戦略  $j$  を選択したときに得られる期待利得を  $u_{ij}$ 、 $\lambda$  をプレイヤーが合理的な意思決定をする程度を表すパラメータとして、確率  $p_{ij} = \exp(\lambda \cdot u_{ij}) / \sum_k \exp(\lambda \cdot u_{ik})$  で戦略  $j$  を選択する AQRE モデルを提案した。また、Zauner [35] は、正規分布に従う意思決定におけるノイズの要素を追加した確率的選択モデルである IPP モデルを提案した。意思決定におけるノイズの程度は正規分布の分散により表現され、これを最尤法により推定している。これらは、ゲームの利得に基づき選択確率を決定するモデルであり、静的確率選択モデルとよばれる。静的確率選択モデルにより計算される選択確率は、被験者の戦略の選択確率の平均値とほぼ一致するが、選択確率の時間的変化、およびセッションごとの結果の違いを説明できない。

Rapoport et al. [27] は、直前のゲームにおける終了ノードに依存して、プレイヤーが選択確率を適応的に学習する PUL (Population Updating Learning) モデル、および IUL (Individual Updating Learning) モデルを提案した。いずれの行動選択モデルも、直前のゲームで *take* が選択されたノードにおける *take* の選択確率は増加し、それよりもノード番号の小さいノードでの選択確率を減少させる確率的選択モデルであり、動的確率選択モデルとよばれる。Rapoport et al. は、動的確率選択モデルのような適応的な行動選択モデルを用いて、3人ムカデゲームの被験者実験における被験者の行動選択分布を再現することに成功したが、すべてのプレイヤーが共通の選択確率に従って行動選択していることから、プレイヤー間の行動の多様性を表現できない。また、動的確率選択モデルでは、行動選択の確率のみで被験者の行動分析を試みていることから、ゲームの利得を考慮しておらず、必ずしも被験者の行動を十分に説明できていないと考えられる。

## 2.3. ゲームモデルに関する適応的なエージェントモデル

Holland and Miller [13] は現実の多くの社会システムは複雑な適応型システムであると考え、これらをモデル化するためには人工適応型エージェントを適用した分析が有効であることを指摘している。Andreoni and Miller [1] は、オークションにおける意思決定を遺伝的アルゴリズムを用いてモデル化し、人工エージェントの意思決定は、オークションに関する実験で観測された人間の意思決定と類似していることを示した。Axelrod [3] は、囚人のジレンマゲームにおいて、人工エージェントの戦略を遺伝的アルゴリズムにより適応的に進化させ

る方法を用い、協調的な戦略選択に関して分析した。Leshno et al. [14] はニューラルネットワークによって意思決定をするエージェントを用いて市場参入ゲームの均衡問題を分析し、Sundali et al. [32] が行った被験者実験の結果と人工適応型エージェントを用いて行ったシミュレーションの結果を比較することで、ニューラルネットワークを通じて意思決定を行う人工適応型エージェントの平均的な意思決定は人間の平均的な意思決定に類似していると報告している。これらの研究を受けて、Nishizaki [24] は、ゲームモデルにおける試行錯誤的な被験者の行動を適応的な学習過程と解釈し、ニューラルネットワークをエージェントの意思決定機構として採用し、エージェントの意思決定に必要と考えられる多様な情報をニューラルネットワークの入力として与え、エージェントの行動を与えられた情報の非線形関数の出力に対応させた。パフォーマンスの悪いエージェント、すなわち、獲得した利得の小さいエージェントを淘汰し、大きい利得を獲得したエージェントを再生させるような遺伝的アルゴリズムを用いてニューラルネットワークのパラメータを改訂するエージェントモデルを提案した。さらに、このエージェントモデルを用いて、協調ゲームにおける被験者の行動分析を行いその有用性を示した。Hayashida et al. [12] は、このエージェントモデルの枠組みを用いて社会的ネットワークに関する分析を行い、累積利得に関するエージェントの長期的視野により、エージェント間に協調関係が構築されることを明らかにした。このように、適応型モデルの有効性はさまざまな研究で確認されており、本論文で開発するシミュレーションシステムの基礎として採用している。

### 3. シミュレーションモデルの開発

#### 3.1. シミュレーションモデルの特徴

前節で述べたように、ムカデゲームの実験では、被験者は部分ゲーム完全均衡から逸脱する行動を取ることが報告されている [22, 27]。Rapoport et al. [27] は2種類の実験を行い、金銭報酬の大きさが被験者の行動に影響を与えることを明らかにした。ムカデゲーム以外のゲームの被験者実験に対しても、均衡概念だけでは説明できない結果がいくつか報告されている [4, 5, 8, 19, 21, 22, 26, 27, 30, 34]。これまでの研究では、均衡からの逸脱の原因として、被験者の (i) 試行錯誤的適応行動 [5, 26]、(ii) 利得に対するリスク回避的な選好 [8, 27]、利得や選好の (iii) 非対称性 [4, 22, 27, 30, 34] による説明が試みられている。また、(iv) 意思決定におけるノイズあるいはエラーを考慮した研究として、Ochs [26] は一般化コイン一致ゲームにおける被験者の行動に対して、ランダム性を導入したモデルによる説明を試みた。この結果に対して McKelvey and Palfrey [19] は高い期待利得をもつ行動は選択されやすいと解釈し、意思決定にノイズあるいはエラーを含むことを考慮した均衡概念を提案している [8]。これらの結果を参考にすれば、ムカデゲームにおいても意思決定におけるノイズを考慮した分析が有用であるといえ、われわれは適応型エージェントをもつシミュレーションモデルを開発し、被験者の行動の説明を試みる。本論文では、ムカデゲームにおける被験者の行動も多様な情報に基づく意思決定を適応的に学習する過程で説明できると考え、人工適応型エージェントを用いたシミュレーションモデルを構築する。

被験者実験 [22, 27] では、被験者は自分のプレイヤー番号や現在のノード番号などの多くの情報に基づいて意思決定していると考えられる。このような複数の情報に基づいた意思決定構造は、市場参入ゲームに関するエージェントモデル [14] と同様に、多入力多出力のニューラルネットワークを用いることでエージェントに実装できると考えられる。また、被験者実験では同じ利得構造を持つゲームが繰り返し行われた結果、図2に示されるように、



実験結果が時間的に変化しており、このことから、被験者は試行錯誤しながら適応的に行動していると考えられる。ニューラルネットワークを用いたエージェントモデル [14] では、重みと閾値が誤差逆伝播法により更新されるが、教師信号を用いる学習法では (i) 試行錯誤的な意思決定構造を実装することが難しい。一方、オークションでのプレイヤーの行動モデルに関するパラメータ値を遺伝的アルゴリズムを用いて決定する枠組み [1] は、エージェント間の相互作用と試行錯誤的な行動選択を考慮した教師信号を用いない適応的なエージェントモデルといえる。エージェントに多くの情報に基づいた試行錯誤的な適応行動を実装する枠組みとして、Nishizaki [24] は、遺伝的アルゴリズムを用いてニューラルネットワークを進化させる枠組みを提案しており、これまでにいくつかのゲームモデルにおける試行錯誤的なエージェントの意思決定システムに適用され、その有用性が確認されている [10, 11, 25]。本論文では、複数の入力情報に基づく意思決定をエージェントに実装するために、ニューラルネットワークをエージェントの意思決定機構として採用する。また、被験者行動の時間的な変化は適応的な学習過程であると考えられるため、Nishizaki [24] と同様の方式を採用する。

本論文では、ムカデゲームに対する被験者実験 [22, 27] では、上記 (i)–(iv) に加えて、ゲームが繰り返し行われていることに着目し、(v) 累積利得最大化についても考慮する。したがって、本論文では (i)–(v) の要因のいずれかまたはいくつかの複合により説明できるものと考え、複数回のゲームの (v) 累積利得最大化、(iii) 非対称性、および (ii) 利得に対するリスク回避的な選好に焦点を当て、(iv) 意思決定におけるノイズも考慮した (i) 試行錯誤的な適応行動を実装したマルチエージェントシミュレーションシステムを構築し、これらの要素で被験者の行動の説明を試みる。

### 3.2. シミュレーションモデルの概要

本論文のシミュレーションモデルの概要を以下に示す。

STEP 1  $3N$  体のニューラルネットワークによって実装された適応型エージェントからなる初期個体群を生成し、 $t = 1$  とする。

STEP 2 集団からランダムに 3 人ずつ選び、 $N$  個のグループを形成する。

STEP 3 各グループのエージェントに 1, 2, 3 のプレイヤー番号をランダムに割り当てる。

STEP 4 各グループでムカデゲームを行う。

STEP 5 遺伝的アルゴリズムを用いてエージェントの学習を行う。ただし、遺伝的アルゴリズムはルーレット選択、1 点交叉、突然変異、エリート保存則により行う。

STEP 6  $t = T$  ならば終了。そうでなければ、 $t := t + 1$  として STEP 2 にもどる。

### 3.3. エージェントの意思決定および学習

エージェントのニューラルネットワークの入力ノードは 12 個あり、ゲームの情報として、意思決定するエージェントに以下の 5 種類の情報を、ニューラルネットワークの入力情報として与える。

- (1) ノード情報：現在のノード番号が 1, 2, 3 のとき (1, 0, 0)、ノード番号が 4, 5, 6 のとき (0, 1, 0)、ノード番号が 7, 8, 9 のとき (0, 0, 1) を 3 つの入力ノードに与える。
- (2) プレイヤー番号：プレイヤー 1 のとき (1, 0, 0)、プレイヤー 2 のとき (0, 1, 0)、プレイヤー 3 のとき (0, 0, 1) を 3 つの入力ノードに与える。
- (3) 期待効用：*take* を選択したときに対応した利得および、のちに述べる *pass* のときに過去  $s$  回のデータから計算される期待利得を 2 つの入力ノードに与える。

- (4) 過去  $s$  回のゲームで得た効用値の重み付平均：各プレイヤー番号に対応した 3 種類の効用値を入力として与える．
- (5) 過去  $s$  回のゲームの各ノードにおいてエージェントが *take* を選択した割合の重み付平均．

金銭報酬を変動させた被験者実験 E-II [27] では，金銭報酬が小さければ協調的な行動を取りやすいことが観測された．このことから，被験者は利得に対して非線形な選好を持つことが考えられる．一般に金銭に対してはリスク回避的な選好を持つことが知られており [16]，被験者も同様に (ii) 利得に対するリスク回避的な選好を持つと考えられる．被験者実験ではゲームで得た利得に比例した金銭報酬が支払われていたため，本論文ではエージェント  $i$  の利得  $v_i$  に関する効用関数は式 (3.1) で与えられるものとする．

$$u_i(v_i) = r_1 + r_2 \exp(-r_3 v_i). \quad (3.1)$$

図 1 に示されるように，1 回のゲームで得られる利得の最大値は 2560 であることから，

$$u_i(v^{\max}) = u_i(3000) = 1, \quad u_i(v^{\min}) = u(0) = 0$$

とした．本論文では，図 1 に示される利得を持つゲームと，これを 1/100 としたゲームの 2 種類に対する事前実験の結果に基づいて， $u(300) = 0.5$  と仮定する．このとき， $r_1 = 1.000987$ ， $r_2 = -1.000987$ ， $r_3 = 0.002307$  となる．

本論文では，複数回のゲームの (v) 累積利得最大化を考慮するため，過去  $s$  回のゲームでエージェント  $i$  が得た利得を  $\Pi_i = (\pi_i^1, \dots, \pi_i^s)$  とする．直近のゲームの利得のほうがより大きく評価されると考えられるため，忘却率を  $\delta \in (0, 1]$  として，エージェント  $i$  が過去  $s$  回のゲームのうちプレイヤー  $j$  として行ったゲームで得た効用値の重み付平均を

$$U(i, j; s) = \frac{\sum_{\tau=1}^s I_{i,j}^{\tau} (1 - \delta)^{\tau-1} u_i(\pi_i^{\tau})}{\sum_{\tau=1}^s I_{i,j}^{\tau} (1 - \delta)^{\tau-1}} \quad (3.2)$$

により計算し，ニューラルネットワークの入力情報 (5) として  $U(i, j; s)$ ， $j = 1, 2, 3$  を与える．ここで， $\tau$  回前のゲームでエージェント  $i$  にプレイヤー番号  $j$  が割り当てられたとき  $I_{i,j}^{\tau} = 1$ ， $k \neq j$  に対しては  $I_{i,k}^{\tau} = 0$  とする．

ノード 1, 2, 3 を  $r = 1$ ，ノード 4, 5, 6 を  $r = 2$ ，ノード 7, 8, 9 を  $r = 3$  として，過去  $s$  回のゲームの各ノードにおけるエージェント  $i$  の行動を  $x_{i,r}^1, x_{i,r}^2, \dots, x_{i,r}^s$  とする．ただし， $\tau$  回前のゲームで *take* を選択した場合を  $x_{i,r}^{\tau} = 1$ ，*pass* を選択した場合  $x_{i,r}^{\tau} = 0$  とする．過去  $s$  回のゲームの各ノードにおいてエージェントが *take* を選択した割合の重み付平均は，

$$X(i, r; s) = \frac{\sum_{\tau=1}^s I_{i,j}^{\tau} (1 - \delta)^{\tau-1} x_{i,r}^{\tau}}{\sum_{\tau=1}^s I_{i,j}^{\tau} (1 - \delta)^{\tau-1}} \quad (3.3)$$

により計算し，ニューラルネットワークの入力情報 (3) として  $X(i, r; s)$ ， $r = 1, 2, 3$  を与える．ただし， $\sum_{\tau=1}^s I_{i,j}^{\tau} (1 - \delta)^{\tau-1} = 0$  のとき， $U(i, j; s) = X(i, r; s) = 0$  とする．

本論文では，エージェントの (iv) 意思決定におけるノイズを考慮するため，エージェントの行動はニューラルネットワークの出力に基づき，戦略はボルツマン選択により確率的に決定されるものとする．ニューラルネットワークの出力ノードは 2 つあり，それぞれ  $out_t, out_p$



とする． $\lambda$  をボルツマン定数として，エージェントは次式で与えられる確率で *take* を選択されるものとする．

$$p_t = \frac{\exp(out_t/\lambda)}{\exp(out_t/\lambda) + \exp(out_p/\lambda)} \quad (3.4)$$

エージェントは，それぞれのニューラルネットワークの重みと閾値により構成される遺伝子を遺伝的アルゴリズムにより進化させる．初期個体群は各遺伝子座に  $[-1, 1]$  のランダムな値を割り当てることで生成する．なお，エージェント  $i$  が過去  $s$  回のゲームの1回あたりの重み付の平均利得を  $f_i(\Pi_i; s) = \sum_{\tau=1}^s (1-\delta)^{\tau-1} \pi_i^\tau / \sum_{\tau=1}^s (1-\delta)^{\tau-1}$  として，式 (3.1) で与えられる効用関数に基づくエージェント  $i$  の効用値  $u_i(f_i(\Pi_i; s))$  を対応する遺伝子の適合度とする．ただし，次節で述べるように遺伝的アルゴリズムは忘却率  $\delta$  に従う3つのエージェントのグループごとに別々に行う．

### 3.4. 非対称なエージェント

被験者実験 [22, 27] において，被験者は複数回のゲームの (v) 累積利得を最大化するために協調関係を構築していたと解釈することができ，また，セッション間の結果の相違から被験者間の協調行動の取りやすさに対する (iii) 非対称性があると考えられる．本論文では，この非対称性を表現するために，過去の情報に対する忘却率を利用する．この非対称性については，社会的ネットワーク形成に関するシミュレーション [12] でも利用されており，有効性が確認されている．このため本論文では，エージェントごとに異なる忘却率を設定することにより，エージェントの非対称性を表現する過去のゲームの影響の強さ  $\delta$  を考え， $\delta$  が大きい場合には直近のゲームの影響を強く受け，短絡的な行動をとるようになり，逆に， $\delta$  が小さい場合には過去のゲームの結果にも影響されるものとする．本論文では， $\delta$  の値によって3種類のエージェントを考え， $\delta = 0.7$  のエージェントを  $A_S$ ， $\delta = 0.5$  のエージェントを  $A_M$ ， $\delta = 0.2$  のエージェントを  $A_L$  と表して3種類のエージェントを用意する．非対称性の影響を調べるために以下の2種類のエージェントの集団を考える．

短絡的エージェントの多い集団  $G_S$ :  $|A_S| = 200, |A_M| = 50, |A_L| = 50$ .

長期的視野に基づく戦略のエージェントの多い集団  $G_L$ :  $|A_S| = 50, |A_M| = 50, |A_L| = 200$ .

## 4. シミュレーション実験

本論文では，まず図1に示されるゲームの木を用いて，2.1節で述べた被験者実験 E-I に対応したシミュレーションを行い，被験者の意思決定に関する (i)–(v) の要素の複合を考慮したシミュレーションモデルが，被験者実験の行動を説明できることを示す．さらに，2種類のグループ  $G_L, G_S$  を考慮することにより，エージェントの (iii) 非対称性の影響について考察する．また，4.2節では (ii) 利得に対するリスク回避的な選好の影響を，4.3節では (iv) 意思決定におけるノイズの影響を中心に，(iii) 非対称性，(v) 累積利得最大化の影響について考察する．なお，シミュレーション実験でのパラメータは，事前実験において最も被験者実験の結果 [27] と整合するパラメータ値を採用し，表1のように設定する．

### 4.1. シミュレーション結果

シミュレーション S-I の集団  $G_S, G_L$  に対する結果として，横軸をノード，縦軸を各ノードでゲームが終了した割合の500期ごとの平均値としたグラフを図4に示す．

図4からわかるように集団  $G_S$  では部分ゲーム完全均衡であるノード1でゲームが終了する割合が高く，集団  $G_L$  では低い．被験者実験の結果を，各ノードでのゲームの終了割合を平均二乗誤差により，図4に示すシミュレーション結果および動的確率選択モデルである

表 1: シミュレーションのパラメータ設定

シミュレーション期間	2000 期	エージェント数	$3 \times N = 300$
過去のゲームを考慮する期間	$s = 10$	ボルツマン定数	$\lambda = 0.05$
交叉確率	$p_c = 0.7$	突然変異確率	$p_m = 0.003$
リスク回避的な効用関数のパラメータ	$(r_1, r_2, r_3) = (1.000987, -1.000987, 0.00237)$		

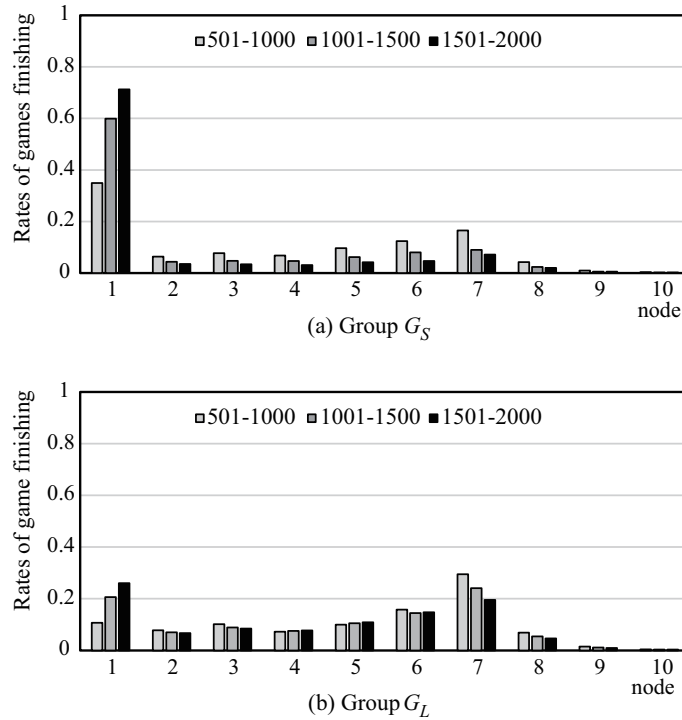


図 4: Result of simulation S-I

PUL, IUL [27] と比較する．図 2 と同様に，ここでは各ノードでのゲームの終了割合の平均値および，時系列変化についての比較を行うため，被験者実験を 1–20 期，21–40 期，41–60 期と 3 分割し，各期間でのゲームの終了割合の平均二乗誤差も比較する．ただし，シミュレーションの 1–500 期を初期学習期間として，501 期目からの 25 期間ごとの平均を被験者実験の 1 期のゲームに対応付けた．すなわち，シミュレーションの 501–1000 期が被験者実験の 1–20 期，1001–1500 期が被験者実験の 21–40 期，1501–2000 期が被験者実験の 41–60 期に対応付けられている．各ノードでのゲームの終了割合の平均二乗誤差を表 2 に示す．

表 2: 被験者実験 [27] でのゲームの終了割合との平均二乗誤差

期数	Experiment E-I-A			Experiment E-I-B		
	PUL	IUL	Simulation	PUL	IUL	Simulation
1–20 期	0.1046	0.1046	<b>0.0143</b>	0.0643	0.0665	<b>0.0109</b>
21–40 期	0.1254	0.1254	<b>0.0162</b>	0.0723	0.0700	<b>0.0162</b>
41–60 期	0.1609	0.1609	<b>0.0058</b>	0.0830	0.0703	<b>0.0186</b>
全期間 (1–60 期)	0.1257	0.1257	<b>0.0125</b>	0.0715	0.0743	<b>0.0159</b>

表2は、各項目で最も小さい平均二乗誤差が太字で示されており、すべての項目において本論文で提案するシミュレーションモデルの誤差が最も小さくなっていることがわかる。動的確率選択モデルPUL, IUL [27]においても、本論文と同様に被験者実験の結果と最も整合するパラメータを選択していることを考えると、動的確率選択モデルと比較して、ニューラルネットワークと遺伝的アルゴリズムを用いたシミュレーションモデルにより、被験者実験 [27] の結果により類似した結果が得られたと言える。

過去のゲームで大きな番号のノードで高い利得を得た経験があれば、 $A_L$  に属するエージェントは忘却率が低いため、*pass* の選択が助長され小さな番号のノードで *take* を選択しにくくなる。このために、直近のゲームが小さな番号のノードで終了しても、 $A_L$  に属するエージェントが多く含まれるために集団  $G_L$  では、図4に示されるようにノード1, 2, 3でゲームが終了する割合が高くならずに協調関係が構築されていたと考えられる。逆に、直近で小さい番号のノードでゲームが終了すれば、 $A_S$  に属するエージェントは *take* を選択しやすくなり、図4に示されるように集団  $G_S$  でノード1でゲームが終了する割合が高くなったと考えられる。

#### 4.2. リスク回避的な選好の影響について

被験者実験 [27] では、金銭報酬の大きさが被験者の行動に影響し、図3に示されるように、金銭利得が1/100になると、ゲームが終了するノード番号が大きくなる傾向が観測された。この効果を検証するために、被験者実験と同様にゲームの利得を1/100とした、集団  $G_S$  に対するシミュレーションを実施した。その結果を図5に示す。図中に示される横軸と縦軸は図4と同様である。比較のために、利得の大きいシミュレーションS-Iの最終500期の結果も示す。

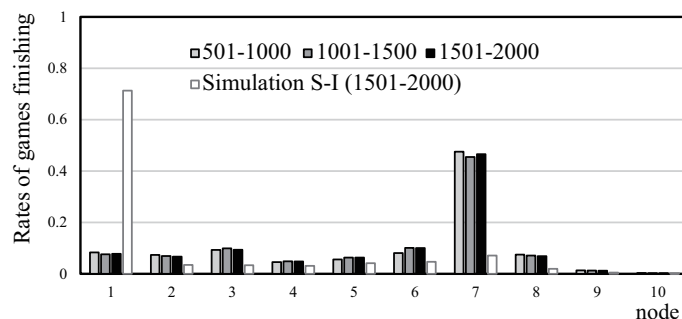


図5: Result of simulation for risk averse preference (Group  $G_S$ )

図5より、シミュレーションS-Iではノード1で終了する割合の高かった集団  $G_S$  でも、金銭報酬の大きさを1/100にすれば被験者実験E-II [27]と同様に、ゲームが終了するノードの番号が大きくなる傾向があることがわかる。この原因は式(3.1)で与えられるエージェントのリスク回避的な選好構造によるものと考えられる。

この効果を明らかにするために、効用関数(3.1)のリスク回避の効果を検証するためのシミュレーションを行う。一般に、金銭に対する選好態度はリスク回避的であることが知られており、エージェントの効用関数  $u_i(v_i)$  が式(3.1)で与えられる場合、確率0.5ずつで  $v_i^{\max} = 3000$ ,  $v_i^{\min} = 0$  を得るくじに対する確実同値  $\hat{v}_i$  を与えることにより  $u_i(v_i)$  の形状が決定される。すなわち、くじに対する確実同値  $\hat{v}_i$  が小さくなればエージェント  $i$  のリスク回避の程度が大きくなり、 $\hat{v}_i$  が大きくなればリスク回避の程度が小さくなる。 $\hat{v}_i$  の値を200, 400,

..., 1400, 1500, 1600, 1800, ..., 2800 と変動させ, 利得をシミュレーション S-I と同様にしたシミュレーション Risk-S-I と, 利得を  $1/100$  としたシミュレーション Risk-S-II の結果を図 6 に示す. ただし, 横軸は確実同値  $\hat{v}_i$ , 縦軸はゲームの終了ノード番号の 100 回の試行の平均である. また, 確実同値に関して,  $\hat{v}_i > (v_i^{\max} + v_i^{\min})$  が成り立つ場合, 利得に対する選好はリスク受容的となる. 図 6 に示される破線は, リスクに対して回避的となる場合と受容的となる場合との境界を示す.

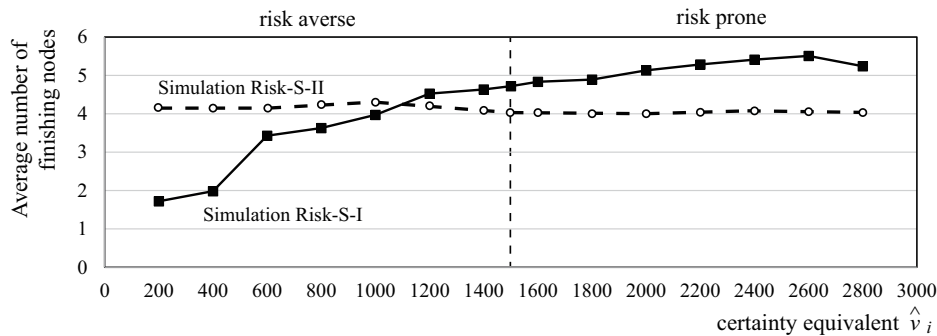


図 6: Result of simulations Risk-S-I and Risk-S-II (Group  $G_S$ )

図 6 より, 確実同値額  $\hat{v}_i$  が大きくなり利得に対するリスク回避の程度が小さくなると, シミュレーション Risk-S-I ではゲームの終了ノードの番号が大きくなり協調関係が構築されるが, シミュレーション Risk-S-II では  $\hat{v}_i$  の大きさにかかわらず協調関係が構築されることがわかる. シミュレーション Risk-S-I ではゲームの結果により大きな利得が得られるため, 自分が大きな番号のノードで *take* を選択する場合と比べると, 他のプレイヤーが小さい番号で *take* を選択した場合に失う利得がシミュレーション Risk-S-II よりも大きい. このことから, 利得に対するエージェントのリスク回避の程度が大きくなれば, リスクの少ない意思決定をするようになり, 図 6 に示されるように小さい番号のノードで *take* を選択し, 協調関係の構築に失敗したものと考えられる.

ノード 1 およびノード 9 で *take* を選択したときの効用の比率は, 利得の大きさの異なるシミュレーション Risk-S-I, Risk-S-II では異なる. 例えば, シミュレーション Risk-S-I と Risk-S-II で結果にかなり差異がある  $\hat{v}_i = 300$  のとき, シミュレーション Risk-S-I では  $u_i(10)/u_i(2560) = 0.0229$  であるのに対して, シミュレーション Risk-S-II では  $u_i(0.01)/u_i(25.6) = 0.00402$  と小さい. このことから, シミュレーション Risk-S-II では, Risk-S-I に比べて小さい番号のノードで *take* を積極的に選択しにくくなり, ゲームの終了ノードの番号が大きくなったものと考えられる. このような利得に対するリスク回避的な選好構造により, シミュレーション S-I に比べて 1 回のゲームでの利得が小さい S-II では集団  $G_S$  の場合でもエージェント間に協調関係が構築されたものと考えられる.

以上の結果から, 利得の大きさによって異なる被験者の行動は (ii) 利得に対するリスク回避的な選好が影響していると考えられる.

#### 4.3. 意思決定におけるノイズの影響について

行動選択確率 (3.4) におけるパラメータ  $\lambda$  はエージェントの意思決定におけるノイズの度合を表現していると解釈でき, また, 別の観点からは, より良い結果を探索するための実験的行動と解釈できる. ここでは, ノイズの影響を分析するために, シミュレーション S-I と同様の設定で, ノイズパラメータを  $\lambda = 0.01, 0.03, 0.05, 0.07, 0.09, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0$

と変動させたシミュレーション S-Error を行う．意思決定におけるノイズは，直近のゲームの結果の影響が大きい短絡的な意思決定をする  $\delta = 0.7$  のエージェントに対して強く影響すると考えられる．このため，シミュレーション S-Error では短絡的に意思決定するエージェントの多い集団  $G_S$  を用いる．シミュレーションの結果として，横軸を  $\lambda$ ，縦軸をゲームが終了したノード番号の 100 回の試行の平均値を，図 7 に示す．

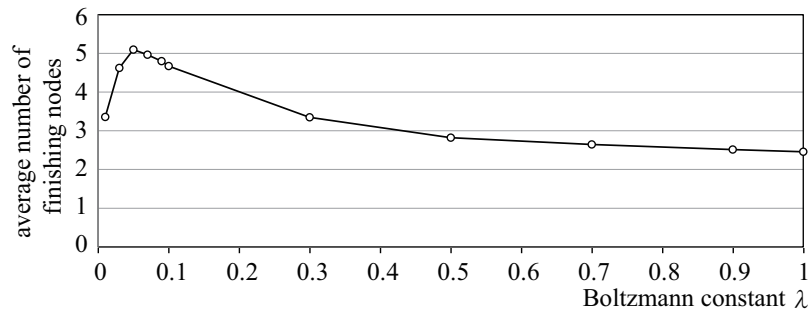


図 7: Result of simulation S-Error (Group  $G_S$ )

式 (3.4) より， $\lambda$  が大きくなれば意思決定におけるノイズの影響が大きくなる．図 7 より， $\lambda = 0.1$  までは  $\lambda$  が大きくなれば終了ノード番号が約 5 まで大きくなり，それ以上  $\lambda$  が大きくなると終了ノード番号が小さくなる．このことから，意思決定におけるノイズが小さい場合，エージェントが *pass* を連続で選択して大きな番号のノードでゲームが終了すれば，そのゲームの結果が他のエージェントの意思決定に影響して協調的な行動を選択するようになると考えられる． $0 < \lambda \leq 0.1$  の範囲では， $\lambda$  が大きくなると実験的探索の効果が高まり，協調が成功する割合が増加する．すなわち， $\lambda = 0.1$  を超えない範囲のノイズあるいは実験的探索は協調を促進するといえる．また， $\lambda > 0.1$  の範囲では，意思決定におけるエラーを起こす割合が高く，エージェントの適切な学習が難しくなるため，図 7 に示されるような結果となったものと考えられる．

被験者実験 E-I においてセッションによって 2 種類の結果が得られたが，シミュレーションで示されるように (v) 累積利得最大化の観点から短絡的な戦略変更をする被験者と長期的視野に基づき戦略を決定する被験者が存在し，集団によってその割合が異なり (iii) 非対称性に違いがあるため，特徴の異なる結果が得られたと考えられる．さらに，図 7 に示されるシミュレーション S-Error の結果から，(iv) 意思決定におけるノイズも協調関係の構築に関連していると考えられる．以上のシミュレーションの結果から被験者間に協調関係が構築された理由として，(v) 累積利得最大化，(iii) 非対称性，(iv) 意思決定におけるノイズの複合が影響していると考えられる．

## 5. おわりに

本論文では，Rapoport et al. [27] の 3 人ムカデゲームの被験者実験に対応するシミュレーション分析を行った．ムカデゲームを含めた多くのゲームモデルにおける被験者の行動は均衡概念のみによる説明は難しく，さまざまな要因に基づく説明が試みられている．本論文では，(i) 試行錯誤的適応行動，(ii) 利得に対するリスク回避的な選好，(iii) 非対称性，(iv) 意思決定におけるノイズ，(v) 累積利得最大化の 5 つの要素に着目したシミュレーション分析を行い，これらにより被験者の行動が説明できることを明らかにした．適応的に行動するエージェントは，ニューラルネットワークと遺伝的アルゴリズムに基づいて設計され，シ



ミュレーションモデルはリスク回避的な効用関数，忘却率を考慮した適合度関数，確率的な意思決定などの枠組みが含まれる．さらに，意思決定におけるノイズや利得に対するリスク回避の影響についての分析から，上記5つの要素が被験者の行動に関連していることを示した．

様々なゲームモデルに対する被験者実験では同じゲームが繰り返し行われる過程で，被験者はゲームの利得構造や，他の被験者との相互作用関係を試行錯誤的に学習していると考えられる．本論文ではこのことに着目し，ニューラルネットワークと遺伝的アルゴリズムに基づく人工適応型エージェントを用いて，ムカデゲームに対するエージェントベースシミュレーション分析を行った．同様の行動分析は最終提案ゲーム，協調ゲーム，社会規範形成，社会的ネットワーク形成などのゲームモデルに対しても行われ，その有用性が示されており [10–12, 24, 25]，展開型ゲームを含めた多くのゲームモデルにおける行動分析に対して有用であると考えられる．特に，市場参入ゲームやオークションゲームなどに対する適用も今後の課題として考えられる．

## 参考文献

- [1] J. Andreoni and J.H. Miller: Auctions with artificial adaptive agents. *Games and Economic Behavior*, **10** (1995), 39–64.
- [2] R.J. Aumann: On the centipede game. *Games and Economic Behavior*, **23** (1998), 91–105.
- [3] R. Axelrod: *The Complexity of Cooperation* (Princeton University Press, Princeton, 1997), 14–29.
- [4] S.K. Berninghaus, K.M. Ehrhart, M. Ott and B. Vogt: Evolution of networks - an experimental analysis. *Journal of Evolutionary Economics*, **17** (2007), 317–347.
- [5] I. Erev and A.E. Roth: Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Games and Economic Behavior*, **23** (1998), 146–175.
- [6] E. Fehr and K. M. Palfrey: A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, **114** (1999), 817–868.
- [7] M. Fey, R.D. McKelvey and T.R. Palfrey: An experimental study of constant-sum centipede games. *International Journal of Game Theory*, **25** (1996), 269–287.
- [8] J.K. Goeree, C.A. Holt and T.R. Palfrey: Risk averse behavior in generalized matching pennies games. *Games and Economic Behavior*, **45** (2003), 97–113.
- [9] J. Harsanyi: Games with incomplete information played by Bayesian players; Part I, II, III. *Management Science*, **14** (1967–68), 468–502.
- [10] T. Hayashida, I. Nishizaki and H. Katagiri: Artificial adaptive agent model characterized by learning and fairness in the ultimatum games. *Journal of Telecommunications and Information Technology*, **2007/4** (2007), 36–44.
- [11] T. Hayashida, I. Nishizaki and H. Katagiri: Network formation and social reputation: a theoretical model and simulation analysis. *International Journal of Knowledge Engineering and Soft Data Paradigms*, **2** (2010), 349–377.

- [12] T. Hayashida, I. Nishizaki, H. Katagiri and R. Kambara: Behavioral analysis in network formation using agent-based simulation systems. *International Journal of Knowledge Engineering and Soft Data Paradigms*, **3** (2011), 22–39.
- [13] J.H. Holland and J.H. Miller: Adaptive intelligent agents in economic theory. *American Economic Review*, **81** (1991), 365–370.
- [14] M. Leshno, D. Moller and P. Ein-Dor: Neural nets in a group decision process. *International Journal of Game Theory*, **31** (2002), 447–467.
- [15] T. Kawagoe and H. Takizawa: Level- $k$  analysis of experimental centipede games. *Journal of Economic Behavior and Organization*, **82** (2012), 548–566.
- [16] R.L. Keeney and H. Raiffa: *Decision with Multiple Objectives: Preferences and Value Tradeoffs* (Wiley, New York, 1976).
- [17] C.R. Mcintosh, J.F. Shogren and A.J. Moravec: Can tournaments induce rational play in the centipede game? Exploring dominance vs. strategic uncertainty. *Economics Bulletin*, **29** (2009), 115–141.
- [18] R.D. McKelvey and T.H. Palfrey: An experimental study of the centipede game. *Econometrica*, **60** (1992), 803–836.
- [19] R.D. McKelvey and T.R. Palfrey: Quantal response equilibria for normal form games. *Games and Economic Behavior*, **10** (1995), 6–38.
- [20] R.D. McKelvey and T.H. Palfrey: Quantal response equilibria for extensive form games. *Experimental Economics*, **1** (1998), 9–41.
- [21] R.D. McKelvey, T.R. Palfrey and R.A. Weber: The effects of payoff magnitude and heterogeneity on behavior in  $2 \times 2$  games with unique mixed strategy equilibria. *Journal of Economic Behavior & Organization*, **42** (2000), 523–548.
- [22] R.O. Murphy, A. Rapoport and J.E. Parco: Population learning of cooperative behavior in a three-person centipede. *Rationality and Society*, **16** (2004), 91–120.
- [23] R. Nagel and F.F. Tang: Experimental results on the centipede game in normal form: an investigation of learning. *Journal Mathematical Psychology*, **42** (1998), 356–384.
- [24] I. Nishizaki: A general framework of agent-based simulation for analyzing behavior of players in games. *Journal of Telecommunications and Information Technology*, **2007/4** (2007), 28–35.
- [25] I. Nishizaki, H. Katagiri and T. Oyama: Simulation analysis using multi-agent systems for social norms. *Computational Economics*, **34** (2009), 37–65.
- [26] J. Ochs: Games with unique, mixed strategy equilibria: an experimental study. *Games and Economic Behavior*, **10** (1995), 202–217.
- [27] A. Rapoport, W.E. Stein, J.E. Parco and T.E. Nicholas: Equilibrium play and adaptive learning in a three-person centipede game. *Games and Economic Behavior*, **43** (2003), 239–265.
- [28] R. Rosenthal: Games of perfect information, predatory pricing, and the chain-store paradox. *Journal of Economic Theory*, **25** (1992), 92–100.
- [29] A.E. Roth and I. Erev: Learning in extensive-form games: Experimental data and



- simple dynamic models in the intermediate term. *Games and Economic Behavior*, **8** (1995), 164–212.
- [30] A. Roth, V. Prasnikar, M. Okuno-Fujiwara, S. Zamir: Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. *American Economic Review*, **81** (1991), 1068–1095.
- [31] A.E. Roth, R.L. Slonim, G. Barron and I. Erev: Learning and equilibrium as useful approximations: accuracy of prediction on randomly selected constant sum games. *Economic Theory, special issue: Behavioral Game Theory Symposium*, **33** (2007), 29–51.
- [32] J.A. Sundali, A. Rapoport and D.A. Seale: Coordination in market entry games with symmetric players. *Organizational Behavior and Human Decision Processes*, **64** (1995), 203–218.
- [33] R.S. Sutton and A.G. Barto: *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 1998).
- [34] J.B. Van Huyck, R.C. Battalio and R.O. Beil: Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, **80** (1990), 234–248.
- [35] K.G. Zauner: A payoff uncertainty explanation of results in experimental centipede games. *Games and Economic Behavior*, **26** (1990), 157–185.

林田智弘

広島大学 大学院工学研究院

〒739-8527 広島県東広島市鏡山 1-4-1

E-mail: hayashida@hiroshima-u.ac.jp

## ABSTRACT

DECISION AND LEARNING OF THE HUMAN SUBJECTS  
IN THE CENTIPEDE GAMES USING AGENT-BASED SIMULATION

Tomohiro Hayashida Ichiro Nishizaki Yuya Sugeo  
*Hiroshima University*

A subgame perfect equilibrium is known as that predicts human behavior in the games in extensive form with a high degree of accuracy. However, some experimental results of the centipede game, one of the games in extensive form, such that human subjects deviate from the equilibrium behavior are reported. It is assumed the players are rational in the sense that payoff maximization in the equilibrium theory. From the fact that the games are repeated in the laboratory experiments, it can be thought that the human subjects make decisions not for maximizing the payoff on a single game, but cumulative payoff which the subjects obtain in multiple games in the experiments. This paper shows that the long-term view for cumulative payoff maximization, asymmetry property of subjects, and the risk attitude for the payoffs can explain the behavior of the subjects in the centipede games through simulation using artificial adaptive agents.