インタラクションによるエージェントの行動変化: 複雑さと固定化速度の可視化

Creation of a limited rational self-determining agent for interaction continuation

福田健翔 [1]　森田純哉 [1]
Kento Fukuda[1]　Junya Morita[1]

[1] 静岡大学
[1] Shizuoka University

Abstract: Change is essential in the emergence of agent behavior. Lack of change leads to behavior fixation and boredom. The most effective means of inducing change is interaction with the other party. In interaction, the situation changes in a complex way depending on the behavior of the other party, and various behaviors are generated. In this study, we examined the variation in single and multi-agent behaviors and their fixation rates using the task of inducing complex interactions in a tag maze game.

1　はじめに

エージェントの行動創発において変化は不可欠である．変化の不足は行動の固定化，すなわち飽きを引き起こす．そしてその変化を引き起こすための有効手段がインタラクションである．

近年，人間とインタラクション可能なエージェントの開発が進んでいるが，人間とエージェントの自律性には大きな違いがある．エージェントは人間のような柔軟性を持たないため，インタラクションが継続しづらく，人間とエージェントとの間でのインタラクションが途切れることがある．では，人間同士のインタラクションと人間とエージェントとのインタラクションにはどのような違いがあるのか．

人間同士のインタラクションでは，行動の探り合いが頻繁に起き，他者の行動を推測しながらインタラクションが進んでいる．この他者の行動を読むプロセスは，それぞれの人間が保持する「他者モデル」を介して行われている [1]．他者モデルは他者の行動を予測するための概念構造であり，これによって人間同士のインタラクションは継続的に進んでいる．このような背景から，人間とエージェントのインタラクションを人間同士のインタラクションに近づけるためには，エージェントの行動を予測できるようにする必要がある．人間が持つ限定合理性 [2] を考えると，エージェントが同様に限定合理的な意思決定を行えれば，インタラクションを継続する手がかりとなる可能性がある．

池上 [3] は，行動の複雑化にはインタラクションが必要だと述べていた．これを踏まえ本研究では，特にマルチエージェントの学習に焦点を当てた．マルチエージェントの学習では，他のエージェントの行動を推測する必要があり，これによって新しい行動・戦術が創発し，行動が変容していく．また，行動の中で自分なりの満足化の基準を設定する必要があり，ここに限定合理的な意思決定が生まれる．ただし，マルチエージェントの学習は既に多くの研究で行われている．そのため本研究では「鬼ごっこ迷路ゲーム」に焦点を当てた．この複雑なタスクや環境において，満足化の基準設定に関する限定合理的な思考が顕著になると仮定し，この条件下で他エージェントの予測・推測が必要なマルチエージェントの学習を行うことで，より人間のインタラクションに近づき，人間の意思決定モデルを理解する手がかりを得られるのではないかと考えた．

本研究では，人間とのインタラクションをより自然に行える自律学習エージェントを検討するために，インタラクションを誘発するための課題である鬼ごっこ迷路ゲームでのエージェント学習によってインタラクションをモデル化する．そしてその学習過程で発生するであろう人間のような行動，戦術の創発や限定合理的な意思決定，飽きの指標である行動の固定化についてシングルエージェント学習との比較をエントロピーの観点から比較を行い評価する．

本稿の構成は以下のとおりである．2 章で本研究をの先行研究を含め，関連研究について紹介し，3 章で今回行う課題の説明，4 章でシミュレーションの説明と結果をまとめ，最後に 5 章で本研究のまとめとして結

論を述べるという流れになっている．

2　関連研究

本研究が継続的なインタラクションの意思決定モデルを検討する際に参考にする先行研究は，岡ら [4] による研究と郷田ら [5] による研究である．これらの先行研究では，共通の実験環境である「鬼ごっこ迷路ゲーム」を使用し，個人特性がインタラクションの継続に与える影響に焦点を当てている．具体的には，鬼役と非鬼役に分かれてペアでゲームに取り組む実験参加者が，飽きた際に任意のタイミングでゲームを終了できるような設定で実施された．

岡ら [4] は,「鬼ごっこ迷路ゲーム」を通して個人特性がインタラクションの継続に与える影響を分析した．「鬼ごっこ迷路ゲーム」は，鬼役と非鬼役に分かれてペアでゲームに取り組む実験参加者が，飽きた際に任意のタイミングでゲームを終了できるよう設定されている．共感指数（EQ），システム化指数（SQ），自閉症スペクトラム指数（AQ）とインタラクションの継続の関係が検証され，特に AQ の差がゲームの継続と有意な相関を示したことが報告されている．これは，異なる個人特性が相手の予測と異なる行動によって飽きが抑制され，インタラクションが継続する可能性があることを示唆している．

郷田ら [5] は，岡らの研究を発展させ，個人特性の高低によってペアを細分化して実験を行った．その結果，AQ だけでなく SQ の差もインタラクションの継続に影響を与えることが示唆された．ただし，これらの研究では実験データ数が少なく，インタラクションの継続と個人特性の関係について議論するにはさらなるデータが必要とされている．そして，エージェントでのインタラクションを考える場合，人間の個人特性をどのように実装するかを議論する必要がある．

また，岡らおよび郷田らの研究で用いられた迷路の探索をシミュレーションした研究として，長島ら [6] の研究がある．長島らによる研究では，エージェントと人間のインタラクションを継続させるための要素として，内発的動機づけに焦点を当てている．この研究では，パターンマッチングを通じて現在の状況を過去の経験に結びつけ，これが人間の知的好奇心と対応する可能性があるという仮説を検討した．その結果，パターンマッチングの成功による楽しさのモデル化が環境学習に有効であり，課題の継続を促進することが確認された．

長友 [7] らの研究は，強化学習においてエージェントが戦術を獲得するプロセスを探究することを目的としている．長友らはエージェントがタスクを遂行する際に同様の方策を繰り返し採用することを「戦術の獲得」と定義した．長友らはエージェントの戦術を「思考の癖」と捉え，これを人間の例えに置き換えて「道に迷った時に突き当たりで右折と左折のどちらを選ぶか」といった意識的かつ無意識的な思考として捉えた．この研究内では ML-Agents にサンプルとして初期実装されている SoccerTwos というサッカー型の環境を用い，エージェントの行動変容と戦術獲得をエージェントの移動軌跡を解析することで検証した．結果として，軌跡の解析からエージェントの学習に伴う行動変容と戦術の獲得が確認できることを示唆した．しかし同時に軌跡情報による分析の限界を指摘し，その他様々な特徴量によるエージェントの行動分析がより正確な戦術の可視化に寄与できる可能性に言及した．

3　課題

本研究では，人間の意思決定モデルを検討するための手法として「鬼ごっこ迷路ゲーム」環境での unity-ml-agents[8] による強化学習を提案する．

3.1　システム設計

この研究では，強化学習の実験環境として，岡ら [4] が提案した「鬼ごっこ迷路ゲーム」を採用した．このゲームは，unity を使用して構築され，鬼ごっこと迷路探索が組み合わさった独自のゲームである．エージェントは鬼と逃げ役に分かれ，鬼役は逃げ役を捕まえるために追いかけ，逃げ役は鬼役から逃げながら迷路を探索し，ゴールを目指す．同時に，鬼役は逃げ役のゴール到達を阻止しなければならない．鬼ごっこの勝利には他者の行動の推測が重要であり，そのためには適切な他者モデルを構築する必要がある．そこに迷路探索の要素が加わることでゲームの目的構造が複雑化し，それにともなって各エージェントが満足化の基準を設定する際に限定合理的な選択が求められる．また，マルチエージェントで学習することによって，シングルエージェントの際には固定されていた満足化の基準がインタラクションによってエージェント自身でその都度決定する必要が生じ，そこに限定合理的な意思決定が生まれて多様な行動，戦術が創発すると考えられる．実際のゲームマップは図 1 に示されており，両エージェントは自由にマップを探索できるが，道幅が 1 人分しかなく，すれ違うことはできない．

3.2　基本ルール

本システムは 2 体のエージェントが鬼役と逃げ役に分かれて迷路内で強化学習を行う．エージェントの色

図 1: ゲームマップ

表 1: 報酬条件

条件	鬼役	逃げ役
逃げ役がゴールに到達した時	-1	+1
鬼役が逃げ役を捕まえた時	+1	-1
時間切れの時	+0	+0

は逃げ役の色が白，鬼役の色は赤となっている．役割ごとに得ることができる報酬とその条件を表 1 に示す．逃げ役は図 1 のマップ左上の位置をスタート地点としてマップ右下の青で表示されているゴールを目指す．対して鬼役のスタート地点は逃げ役が目指すゴールの目の前となっている．ゲームは 1 エピソード最大 5000 ステップまで行われ，鬼役の勝利条件は相手を捕まえることで，逃げ役にとっての勝利条件はゴールに到達することである．ゲームに勝利したエージェントは報酬を獲得できる．また，1 エピソードの制限時間である 5000 ステップを経過しても勝敗がついていなかった場合，エピソードはそこで終了し引き分けとなる．

4　シミュレーション

本シミュレーションの目的はシングルエージェントとマルチエージェントでどのように行動が変化するかを検討する目的である．また，その変化した行動のモデルがどのような特徴を持つのか確認する．行動の固定化は飽きによって発生し，飽きは満足化の性質に従う．そしてシングルエージェントでは状況が一定なため満足化の基準が変化しないが，マルチエージェントではその時の状況に応じて自分で満足化の基準を設定を必要があるため限定合理的な思考が生まれると予想した．本シミュレーションでは，行動の変化を可視化するためにシングル，マルチの両方において，エージェントのエピソード終了時の滞在区間と，その滞在区画にどれほど偏りがあるかを示す情報量のエントロピーの合計や時間変化に関する分析を行った．

4.1　シングルエージェント

4.1.1　エージェント設定

このシミュレーションでは，マルチエージェントでの学習に先立ち，鬼役の位置と逃げ役の位置からエージェントを単一で学習させた．鬼役側のエージェントを学習させる際には，逃げ役側の位置を目的地として迷路の探索を行った．逆に逃げ役側のエージェントを学習させる際には通常通りゴールの位置を目的地として迷路探索を行った．どちらの学習も目的地に到達すると正の報酬を+1 受け取り，エピソードが終了し初期位置に戻され，次のエピソードが開始される．1 エピソードの制限時間は 5000 ステップとし，時間内に目的地に到達できなかった場合はエピソードが終了し初期位置に戻り，新しいエピソードが始まる．この時報酬は得られない．本研究では各エージェントに対して 1880 エピソード分の学習を行った．

4.1.2　滞在区画分析

各エージェントのエピソード終了時の滞在区画の分析を行った．エージェントの滞在区画を分析するにあたって，今回 unity で作成した迷路環境で座標の連続値を解釈することが難しかったため，分析を可能にするために先行研究である郷田ら [5] が行ったのと同様にマップを 19 のコースに区画分けして離散化を行った．離散化は大まかにマップの左右どちら側に滞在しているかをわかりやすくするために，縦に数字が連なるようにコースに数字を割り当てた．離散化したマップが図 2 である．

滞在区画の分析では，離散化したマップの区画番号を用いて各エージェントのエピソード終了時の状態を分類し，エピソード数の変化に伴った変化を可視化するために散布図を描いた．作成したものが図 3 である．縦に入っている緑の線は，同じ x 軸にあるコースのまとまりごとに対応させてグラフを区切っている．滞在区画のグラフを観察すると，鬼役側と逃げ役側のエージェントが共に初めの 700 エピソードまでさまざまな区画に滞在していることが分かる．しかし，700 エピソードを過ぎると滞在区画に偏りが生じ，後半ではほぼ同じ区画に滞在する傾向が見られる．鬼役側からスタートしたエージェントは逃げ役側の初期位置である 1 番の区画に，逃げ役側からスタートしたエージェントはゴールが配置されている 19 番の区画に多く滞在していることから，700 エピソードあたりでゴールまで

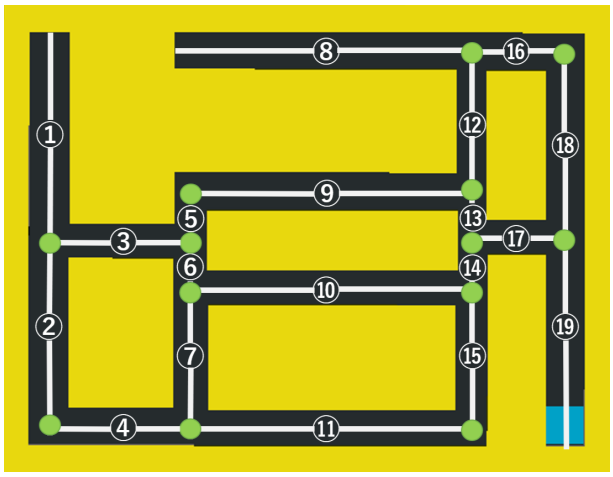

図 2: 離散化マップ

の経路探索がほぼ完成しており，その後は行動が固定化されてゴールへ最速で向かっていること示唆されている．

4.1.3　エントロピー分析

鬼役側のエージェントと逃げ役側のエージェントそれぞれについてのエントロピーの算出を行った．それぞれの区画に滞在している確率を $P(E)$ として以下の式で計算される．

$$I(E) = -logP(E)$$

この情報量に $I(E)$ に $P(E)$ をかけることによって，エントロピーを求める．具体的な計算式は以下の通りである．

$$H = -P(E)logP(E)$$

はじめに，1880 エピソード全体のエントロピーを各エージェントごとに計算した．離散化されたマップを利用して，エピソード終了時の各エージェントの滞在区画を区画ごとに数え，その確率を計算した．得られた確率をエントロピーの式に代入し，全ての区画に対するエントロピーの和を計算した．結果として，鬼役のエントロピーが 2.059，逃げ役のエントロピーが 2.056 となった．左右非対称なマップではあるが行き止まり一つ分しか違いがないためか，エントロピーの差は大きくない．鬼役の方がやや高い傾向が見られるが，これは行き止まりである 8 番の区画への探索のしやすさの違いによるものと推測される．

次にエントロピーの時間変化をグラフ化した．エピソードを 10 ずつに区切り，合計 188 個のエントロピーを算出し，時間変化を可視化した．それが図 4 である．グラフからは，700 エピソードあたりで行動が固定化

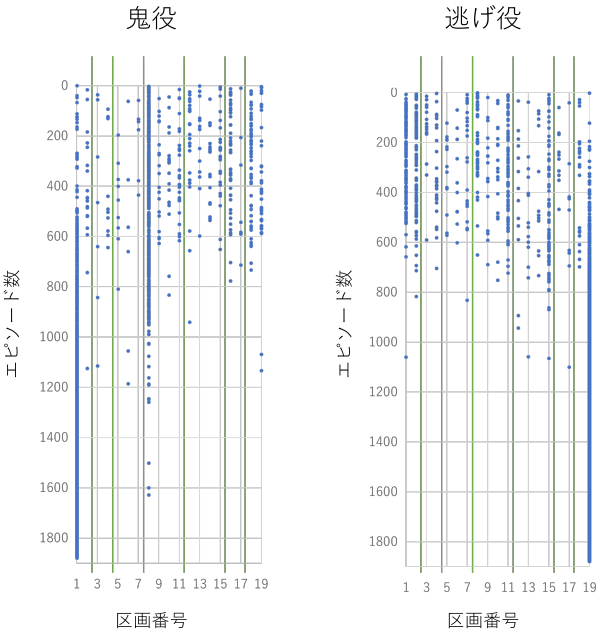

図 3: シングルエージェント滞在区画

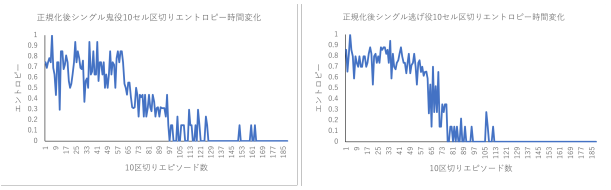

図 4: シングルエージェントエントロピー時間変化

する傾向が明確に示されている．また，両エージェントのエントロピーの段階的な変化から，エージェントの方策の変化が推測できる．600 エピソードあたりで急激なエントロピーの下降が見られ，これはマップを「探索」する方策から「活用」する方策への切り替えを示唆している．その後は目的地までのルートを「最適化」する方策に移行していることがグラフから推察される．

このようにシングルエージェントの学習では，左右非対称であるマップが探索に及ぼす影響や，エージェントがとる方策の時間別変化をマップの離散化とエントロピーの時間変化のグラフ化によって可視化した．

4.2　マルチエージェント

4.2.1　エージェント設定

今回は簡素な報酬で学習を行った．報酬は，表 1 に示した通り，鬼役が逃げ役を捕まえた際に鬼役に+1，逃げ役に−1．逃げ役がゴールに到達した際に逃げ役

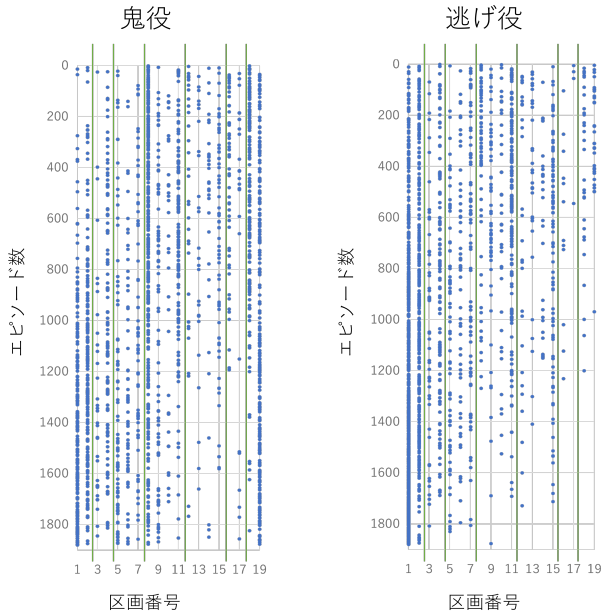

図 5: マルチエージェント滞在区画

に+1，鬼役に−1．引き分けの際は報酬変動はないものとした．

4.2.2 滞在区画分析

シングルエージェントの際と同様に離散化されたマップ内でエピソード終了時の各エージェントの滞在区画を可視化した（図 5）．この図をシングルエージェントのグラフと比較すると，滞在区画の散らばりが著しく大きいことが明らかである．前半では，両エージェントとも滞在区間が散らばり，後半にかけて少し固定化の傾向が見られるが，それでもシングルエージェントと比べて大きな散らばりが維持されている．後半の固定化傾向は，滞在区画の情報が鬼役逃げ役ともに逃げ役のスタート位置に近い道番号である 1 から 6 に偏っていることから，鬼役が優勢だったことが読み取れる．

鬼役が優勢となった原因は，方策の効果的な変化から生まれた戦術にあると推察される．鬼役のエピソード終了時滞在区画を見ると，前半と後半で大きく変わっていることが読み取れる．前半は鬼役のスタート地点に近い 12 から 19 に滞在していることが多かったが，後半になるにつれて逃げ役側のスタート地点の近くである 1 から 6 に滞在することが増えている．対して逃げ役は，前半は滞在区画の散らばりが大きくゴールが配置してある区画である 19 への到達を何度も成功させていたが，後半になるにつれて散らばりは小さくなり，相手エリアでの滞在もほとんどなくなっている．このデータから鬼役と逃げ役の方策の変化が考察できる．鬼役は，はじめは迷路を探索する方策に従っていたものの，逃げ役にゴールに到達されてしまい報酬のマイナスをうけ，これ以上のマイナスをできるだけ減らすために方策を変化させたと思われる．

表 2: エピソード全体のエントロピー

条件	鬼役	逃げ役
シングルエージェント	2.059	2.056
マルチエージェント	3.84	3.31

その方策変化から生まれた戦術として予想されるのがいわゆる「待ち伏せ」である．今回のエージェントの設計上完全に同じ位置に留まることはできないが，大まかなエリアで待ち伏せすることは可能である．後半エピソードになってもゴール目の前の区画である 19 で鬼役が観測されていることからも定期的に待ち伏せの戦術をとっていた確率は高いと言えるだろう．その結果鬼役が報酬を獲得し続ける状況が継続し，逃げ役のエージェントの方策が変化してゴールへ向かうのではなく自分のエリア付近で逃げ回るという戦術に変更せざるを得なくなったと思われる．

後半の両エージェントの滞在区画が逃げ役のスタート地点寄りの区画に偏っているのは，逃げ役がゴールに向かってくることが少なくなり，報酬を得るために逃げ役を捕まえに行く方策をとったからだと考えられる．後半エピソードで鬼役が優勢になっている原因は，待ち伏せ戦術によって前半エピソードで逃げ役に報酬を与えずに鬼役が多くの正の報酬を獲得して効率的に学習を行ったと考えれば辻褄が合うだろう．

4.2.3 エントロピー分析

鬼役と逃げ役のそれぞれについて前セクションと同様の計算方法でエントロピーを算出した．1880 エピソード全体のエントロピーは鬼役が 3.84，逃げ役が 3.31 となった．表 2 はシングルエージェントのエントロピーも含めた結果のまとめである．シングルエージェントの時と同様で鬼役のエントロピーのほうが高いことが確認できるが，今回はその差がシングルエージェントに比べてかなり大きい．さらに，両エージェントのエントロピーの数値もかなり高くなっている．これは先ほどの考察通り，マルチエージェントでの学習によってエージェントの方策が変化し，行動が創発されて戦術が生まれた結果と考えられる．

図 6 は 10 エピソード区切りのエントロピーの時間変化を示す．この図では，シングルエージェントの時に比べて明らかに両エージェントとも高い数値を維持している．これは行動が定期的に変化し続けていて，行動の固定化速度が緩やかであることを示している．さらに鬼役のグラフは途中まで緩やかではあるが一部右

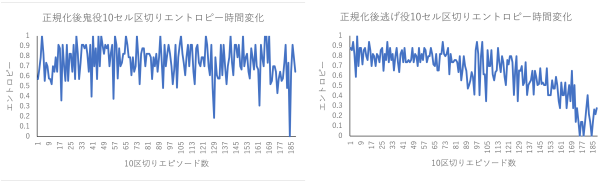

図 6: マルチエージェントエントロピー時間変化

肩上がりの傾向が見られる．また，逃げ役もわずかではあるがエントロピーが上昇し，高い数値を維持しながら緩やかに下がっていっている．同じ行動を繰り返しているだけならばエントロピーは下降していくはずである．これはつまり，エージェントに新たな行動が創発されたことを意味する．

5 結論

本研究では，行動の固定化の変化について確認するために，シングルエージェントの迷路探索ゲームとマルチエージェントの鬼ごっこ迷路ゲームの 2 つの学習を行いエージェントモデルを作成し，その結果を比較した．その結果，シングルエージェントに比べてマルチエージェントでは明らかに滞在位置の偏りが少なく，情報量も全体的に高い水準を示すことが明らかになった．エントロピー時間変化に関しても，シングルエージェントでは行動の固定化が速かったのに対し，マルチエージェントでは行動の固定化までの速度が遅くなっていた．さらにマルチエージェントの鬼役に関しては一部右肩上がりの傾向が見られ行動が固定化されている傾向はみられなかった．

このことから，シングルエージェントでは顕著に表れていた行動の固定化が，マルチエージェントでのインタラクションによって，エージェント間で新たな行動が創発されて戦術が生まれ，行動の固定化を防ぐことを確認できた．また，行動の固定化はタスクに対する満足化の基準と結び付けることができ，マルチエージェントでの学習において逃げ役と鬼役の間で行動の固定化傾向が変化してることから，それぞれのエージェントが独自の満足化の基準を設定したことが読み取れる．そして，その満足化の基準の設定の際に，それぞれのエージェントがお互いの行動を予測しながら限定合理的な意思決定を行ったと推測される．

これらのことから，今回のシミュレーションにおいて，インタラクションの中で相手の動きを推測して戦術を効果的に創発し，満足化の基準を限定合理的に自己決定するエージェントを作成できたと推察される．これはエージェントに戦術，すなわち個人特性が付与されたと考えられる．これによってエージェントの行動を予測しやすくなり，人間とのインタラクションを人間同士のように行うエージェントを検討する上で貢献すると予想される．

今回のシミュレーションでは簡素な報酬しか付与しなかったため，エージェントの目標構造もそれに従い簡素なものとなってしまった．報酬を複雑にすることができればエージェントの目標構造も本来の鬼ごっこ迷路ゲーム同様に多岐にわたり，より多くの行動，戦術の創発がおこると予想される．また，郷田ら [5] の先行研究環境に近づけるために，マップ上への得点アイテムの配置や，一定回数シミュレーションを行った際にエージェントが一定確率でゲームを自ら終了するような実装を行うことができれば，満足化の基準の設定が複雑になり，より詳細な限定合理的な意思決定モデルの検討が可能になると考察する．

参考文献

[1] 横山絢美, 岡田浩之, 大森隆司, 石川悟, 長田悠吾. 自者と他者の双方向行動調節による社会的インタラクションのモデル化. 人工知能学会全国大会論文集 第 21 回 (2007), pp. 2C57–2C57. 一般社団法人 人工知能学会, 2007.

[2] Herbert A. Simon. *The Sciences of the Artificial, third edition (English Edition)*. The MIT Press, 1996. 稲葉元吉, 吉原英樹 (訳), システムの科学 第 3 版, パーソナルメディア, 1999.

[3] 池上高志. 生命理論としての認知科学: 減算と縮約の哲学をめぐって. 認知科学, Vol. 28, No. 2, pp. 198–210, 2021.

[4] 岡真奈美, 森田純哉, 大本義正. 持続的なインタラクションの成立における個人特性の影響˜ 多義的な目標構造を有するゲーム課題を用いた検討˜. 電子情報通信学会技術研究報告; 信学技報.

[5] 郷田怜花, 森田純哉, 大本義正. インタラクションを伴う迷路課題における行動的特徴と個人特性の関係. HAI シンポジウム 2021, 2021.

[6] 長島一真, 森田純哉, 竹内勇剛. Act-r による内発的動機づけのモデル化. 人工知能学会第二種研究会資料, Vol. 2019, No. AGI-013, p. 07, 2019.

[7] 長友結希, 三宅陽一郎ほか. 強化学習によるエージェントの戦術獲得の分析. ゲームプログラミングワークショップ 2021 論文集, Vol. 2021, pp. 106–110, 2021.

[8] Unity. https://unity.com/ja/products/machine-learning-agents.