

# 「楽しさ」と「飽き」に着目したインタラクション継続のモデル化 Modeling in Sustainable Interactions Focusing on “Fun” and “boring”

長島 一真<sup>†</sup>, 森田 純哉<sup>†</sup>, 竹内 勇剛<sup>†</sup>

Kazuma Nagashima, Junya Morita, Yugo Takeuchi

<sup>†</sup> 静岡大学

Shizuoka University

cs16065@s.inf.shizuoka.ac.jp, j-morita@inf.shizuoka.ac.jp, takeuchi@inf.shizuoka.ac.jp

## 概要

近年、人間とインタラクションをするエージェントが身近になってきている。これらのエージェントに対し、多くのユーザは当初はインタラクションに楽しみを見出す。ただし、エージェントの動作を理解するに従い、人間はそのエージェントとのインタラクションに飽きを感じ、いずれ利用をやめてしまう。本研究では、認知アーキテクチャの1つである ACT-R を用い、インタラクションを継続するエージェントの構築を目指す。第一段階として、認知モデルを用いてエージェント間のインタラクションを継続する条件を明らかにすることを試みる。Raphael の “Theory of Fun for Game Design” に基づいて、「楽しさ」と「飽き」のモデリングを行った。エージェントに対して同じ課題を繰り返させ、「面白さ」に基づいたパラメータを変更することで課題の継続数の遷移を確認した。将来的に、このモデルを拡張することにより、「楽しさ」と「飽き」に基づき、エージェント間でインタラクションを継続するシステムの開発を目指す。

キーワード: 認知モデリング, インタラクションの継続

## 1. はじめに

現代の社会において、人とインタラクションするエージェントは偏在している。例えば、複数のゲームユーザに囲まれるオンラインマルチプレイゲームにおける NPC は、ユーザの支援やゲームを通してユーザとコミュニケーションを行う。しかし、これらのエージェントとのインタラクションにおいて、通常、ユーザは人間同士のインタラクションで感じる楽しさを抱くことが少なく、飽きにより利用を停止する問題が生じている。

このような問題は、インタラクションの持続性の問題として以前より研究されている [5]。この研究で指摘されることは、相手の行動の予測可能性がインタラクションの持続に影響することである。ユーザがエー

ジェントの行動を予測可能になったとき、エージェントとのインタラクションをやめてしまう。またこの研究において、持続的なインタラクションを行うエージェントの条件として、ユーザに対する振る舞いを適応的に変化させていくことなどが挙げられている。

本研究では、このようなインタラクションの持続の背景にあるメカニズムを認知モデルの観点から検討する。ここで言う認知モデルとは、コンピュータ上に実装された人間の内部処理に関わる仮説である。通常の人工知能研究に対して、認知モデルの研究は、人間のエラーやバイアスを説明し、予測することに重点を置く。また、認知モデルはしばしば認知アーキテクチャを利用して構築される。認知アーキテクチャとは、個別の課題において、認知モデルを実装するフレームワークである。認知アーキテクチャを利用したモデルの実装により、一般的な認知機能との関連の中で個別の課題に関するモデルを構築できる。これまでに複数の認知アーキテクチャが開発されてきた。本研究ではその中でも大きなコミュニティを持つ ACT-R (Adaptive Control of Thought-Rational [2]) に着目する。

本研究の目的は、ACT-R から提供されるプリミティブな認知プロセスの集積によって、インタラクションの持続がどのように成し遂げられるのかを明らかにすることである。その方法として、実際に人間とエージェントが仮想環境においてインタラクションする課題を想定し、実験において得られた人間の行動特性を再現するモデルを構築する。本研究では、その第一段階として人間を対象とした実験の結果を示し、その結果を再現するインタラクション継続のモデル化を議論する。また、その議論の proof of concept として構築したプロトタイプ的なモデルによる予備的なシミュレーション結果を報告する。

以下のセクションでは、2 節において本研究と関連する研究を示し、本研究の狙いを明確化する。3 節で

はタスクとデータ, 4 節にて想定するインタラクションのモデル, 5 節にてプロトタイプモデルを示す. 本論文の最後に, 現状の到達点をまとめ, 今後の方向を示す.

## 2. 関連研究

### 2.1 内発的動機づけによるインタラクションの継続

インタラクションの継続に関するモデル化は本研究において初めて検討されるものではない. 強化学習の枠組みを利用する複数の研究がインタラクション継続の要因を検討している [15, 14]. これらの研究は, 環境の予測可能性の観点から報酬を決定する Intrinsically motivated reinforcement learning (IMRL) [7] の枠組みの中でインタラクションの継続を説明してきた.

通常, 強化学習におけるエージェントは, 外部環境と継続的にインタラクションを行う. エージェントは, 環境から報酬を受け取り, 時間経過の中で, それを最大化するように努める. 一方, IMRL では, Sutton が述べたエージェントと環境間の境界と, 身体と環境間の物理的な境界は同様ではない [17] という主張を援用し, 環境を外部環境と内部環境に分け報酬を別々に定義する. 外部環境から直接報酬を受ける従来の強化学習に対し, IMRL は, 内部環境の状態によって報酬が変動するものであり, 予期せぬ反応に対する好奇心などをモデル化するものである. このようなメカニズムを導入することで, 従来の強化学習よりも, エージェントの高レベルのスキル獲得課題においてパフォーマンスが向上するとされる.

インタラクション継続の文脈に, 上記の IMRL を導入すれば, 他者に対する驚きを伴う予測可能性がインタラクション継続に影響することが示唆される. 野澤と近藤は IMRL を含むエージェント間のインタラクションが継続的なものになることを, シミュレーションによって示している. また, Quresi らは, 人と継続的なインタラクションを行うロボットに対し, End-to-End [12] の IMRL によって, ロボットに社会的スキルを身につけさせる試みを報告している [15]. しかし, これらの研究において, モデルの状態はマルコフ過程に沿って確率的に遷移するため, 学習途中の状態を制御するのは難しい. また, End-to-End の学習において, 学習の過程はブラックボックス化されるので明らかではない.

### 2.2 ゲーム課題における認知モデルの利用

上記の強化学習に関わる研究は, 人間とモデルの具体的な行動の対応を検討するものではない. 人間の行動との詳細な対応を説明するモデルは, ACT-R を用いて構築できる. ACT-R はプロダクションシステム [13] をベースとした認知アーキテクチャであり, 脳部位と対応するモジュールを持つ. モジュールには, 外界との入出力を制御するモジュール群, 経験や知識を保持する宣言的モジュール, タスクにおける状態を管理するゴールモジュールなどが含まれる. ACT-R におけるプロダクションルールは各モジュールの状態を条件として選択され, モジュールに対するコマンド (例: 画面の特定の位置に注意を向ける, 条件に合う知識を検索する, タスクの状態の認識を更新する) をアクションとして発行する. 1 つのプロダクションの発火に要する時間や各モジュールがそれぞれのコマンドを遂行する時間は, 過去の心理実験の結果によって見積もられている. そのため, ACT-R は特定のタスクの遂行に要する時間を予測することができ, 人間の行動との詳細な対応を可能にする.

インタラクティブ環境における ACT-R によるモデリングとして, ゲーム課題を解く認知モデルの研究が行われている. その中に迷路課題を扱う研究が存在する. Fu らは, ACT-R のモデルに上下左右の知識を与え, 強化学習の手法を用いて迷路課題を繰り返し解かせる研究を行っている [19]. ACT-R は Q-Learning [20] に似たユーティリティモジュールという機能が存在し, このモデルでは正しい行動をした時に正の報酬を, 間違った行動した時に負の報酬を与えている. これによって, 課題の試行回数を増やすことで, モデルが最適な行動を取るよう学習していく. この研究において課題試行回数におけるモデルの実験データと, 過去の動物と人間の実験データが適合していると主張している.

また, 迷路環境における学習としては, 強化学習だけでなく, ACT-R のモデルの宣言的知識を用いたを用いたパスプランニングを行う研究が存在する [16]. この研究において, 位置情報をモデルが利用できるアフオーダンス [9] として宣言的知識に入れ, トポロジカルマップを構成し, バックトラックを用いてパスのプランニングを行っている.

上記一連の研究は環境探索における多様な戦略を明らかにし, 人間の個人差や学習のプロセスを説明する. しかし, これらはタスクの継続をモデル化するものではない. 本研究では, 上記の研究を引き継ぎつつ,

人間の行動と直接対応づける課題において、インタラクション継続のモデル化を狙う。以下にて、本研究において採用する実験課題を示す。

### 3. 課題

本研究では、岡ら [21] によって習得された、インタラクションの継続に関するデータをモデル化の対象とする。この研究において、岡らは人と人がインタラクションするオリジナルの課題を構築し、実験参加者からデータを取得した。

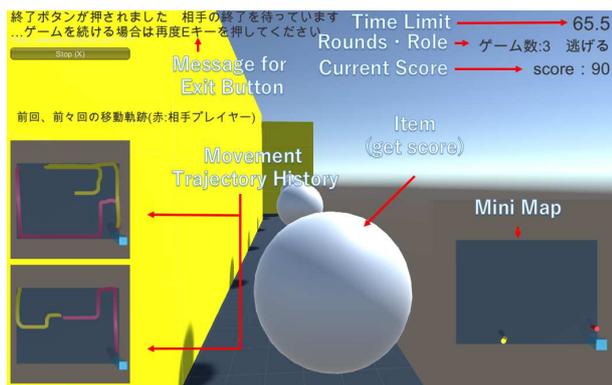


図1 人同士の実験環境

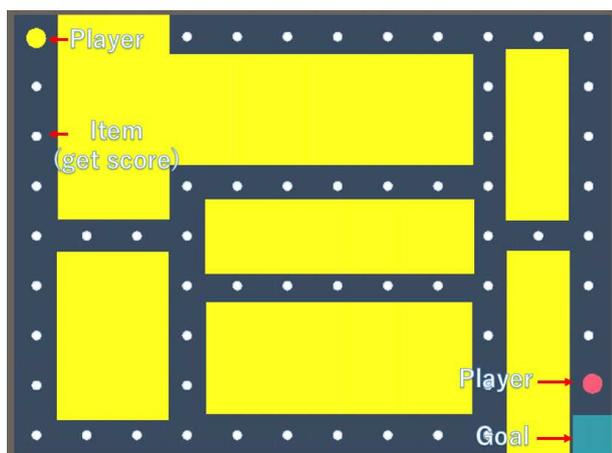


図2 ゲームマップ

岡らが考案したのは、仮想 3D 空間 (図 1) で「鬼ごっこ」と「迷路ゲーム」を組み合わせた「鬼ごっこ迷路ゲーム」課題である。この課題において、迷路内に存在する 2 人のプレイヤーは鬼役と非鬼役を交互に切り替えながらゲームを遂行する。鬼役は非鬼役を追いかけ、追いついたら勝利する。非鬼役は鬼役から逃げ、迷路内に設置された出口にたどり着くと勝利する。ゲームにおけるマップ (図 2) はラウンドにかかわらず固定とする。つまり、ゲームを繰り返すことで鬼役も非鬼役も相手から効果的に逃げ、追うことができるようになる。

本研究はインタラクション継続の要員として、課題に参加するプレイヤー間の個人特性に注目する。どのような特性をもつ個人、あるいはペアがインタラクションを継続しやすいのかを検討する。

具体的な個人特性として、バロンコーエン [4] により構築された自閉症スペクトラム指数 (AQ)、共感指数 (EQ)、システム化指数 (SQ) を採用している。自閉症スペクトラムは、特殊な人のみが有する障害ではなく、一般の人からの連続的なスペクトラムであるとされ、インタラクションの継続に顕著に働く個人特性であると考えられている。また、システム化指数と共感指数は自閉症スペクトラムを構成する個人特性であるとされる。高いシステム化指数と低い共感指数により、自閉傾向の強さを表現できる。これらの指標は、自閉症スペクトラム指数よりも、詳細で一般的な粒度でコミュニケーションに関わる個人特性を取得できると考えられている。実験参加者は、これらの個人特性を測定する質問紙を、ゲーム終了後に回答した。この実験において得られたデータを分析した結果、以下のことが明らかになっている。

- ゲームの成績と質問紙によって測定される個人特性 (AQ/EQ/SQ) の相関を検討した結果、システム化指数とゲームの成績に有意な正の相関が観察された。
- 勝率の高いプレイヤーが先に終了ボタンを押す比率は、勝率の低いプレイヤーが先に終了ボタンを押す比率よりも小さかった。しかし、その傾向は統計的に有意とはならなかった。
- 個人特性の組み合わせ (ペア内での個人特性の合計と差分) とインタラクションの継続 (一方が終了ボタンを押したラウンド、両方が終了ボタンを押したラウンド) の相関を検討したところ、自閉症スペクトラム指数の差分とインタラクション継続に正の相関が観察された

実験結果が示すのは、ゲーム中の振る舞いに個人差が存在すること、ペア内での個人特性の組み合わせがインタラクション継続に関係する可能性があることである。

このうち、ゲーム内での振る舞いの個人差は、ACT-R によるパラメータの変更によって再現できる可能性がある。近年の ACT-R コミュニティでは、様々な精神疾患を説明する計算精神医学的なモデルの研究が進められている [18]。

また、ゲームの勝敗がインタラクションの継続と強く関連しなかった結果については、ゲームに勝つことによる正の報酬とゲームの習熟による予測可能性の

間のトレードオフが生じたと解釈できる。ゲームに勝つためには、ゲームに習熟する必要がある。しかし、ゲームに習熟すると予測可能性が高まり、飽きを生じさせると考えられる。

異なる個人特性を持つ相手とのインタラクションにおいて、課題が継続した結果は、対戦相手の振る舞いの予測可能性の観点から解釈できる。通常、自分とは異なる行動を行う他者に対しては、自分の振る舞いに基づき予測を行うことが困難であると考えられる。高いAQを持つ人は、他の高いAQを持つ人々に対して優れた推論能力を持つという仮説が、過去の心理学の研究によって裏付けされている [10]。

また、SQの差ではなく、AQの差において関連したことも興味深い結果である。SQ差が相関しなかったことは、SQがゲーム成績と強く関連する個人特性であることと関連する可能性がある。それに対して、AQはゲーム成績と直接関連する個人特性ではなく、純粋な予測可能性のみに関連した可能性がある。

## 4. 課題継続の ACT-R モデル

### 4.1 ACT-R の学習理論

上記の実験において得られた知見を説明するモデルを ACT-R のメカニズムを用いることで考える。その際、本研究では、内発的動機の生起要因の1つとして、楽しさに注目する。楽しさに関する代表的な理論としては、Raphael による “Theory of Fun for Game Design[11]” が考えられる。この理論では、ゲームにおける「楽しさ」をパターンを発見し、学習を続ける事であると考えられる。例えば、数あるパターンから最適解が発見されたゲームは、最適解発見後にゲームから得られるものがなくなり、「飽き」が生じてしまう。

我々は、ACT-R の「プロダクションコンパイルモジュール」と「ユーティリティモジュール」を用いて「楽しさ」と「飽き」を表現することができると考えている。以下、これらの機能の概略を説明する。詳細は文献 [2, 6] あるいは ACT-R のチュートリアル (unit 6,7) などを参照されたい。

#### 4.1.1 プロダクションコンパイル

「プロダクションコンパイル」とは、2つの定義されているプロダクションルールを1つのプロダクションルールに統合する機能である [3]。ある課題に対しての一連のルールを反復発火することで、ルールの統合が起き、課題達成までに発火するルールの数が減る。ルール統合の対象となるルールは、通常ルールの条件節に変数を含むものとなる。ACT-R において、記憶

に基づく柔軟な手続きの遂行過程は、ルールの条件節に記述される変数のパターンと宣言的知識（記憶）のマッチングによってモデル化される。それに対して、プロダクションコンパイルによるルールの統合は、そのようなパターンマッチによる宣言的知識の検索をスキップする。つまり、統合前のルールに含まれていた変数は、個別の宣言的知識の値によって置換される。そのため、定型的で自動的な課題遂行の手続きがシミュレーションされるようになる。

図3は鬼ごっこゲームにおいて、相手の予測位置からパスのプランニングをするプロセスを ACT-R のトレースとして示している。縦軸に時間を示し、各コラムはモジュールのイベントを示している。2.2 で述べたように ACT-R は複数のモジュールを持つが、ここではプロダクションモジュールと宣言的モジュールのみを示している。この ACT-R モデルは、宣言的知識に相手の情報、探索方向、環境情報が保持されている。モデルは、このプロセス中に宣言的知識から記憶検索を行い、相手の予測位置までのパスを求める。ACT-R のパラメータが標準の状態では、1つのルールの実行に 50msec の時間コストがかかる。また、宣言的モジュールからの宣言的知識の検索についても、その記憶の活性値に応じた時間を要する。そのため、1つの課題を行う際に、宣言的記憶の検索回数が少ないほど思考する時間が減少する。

本モジュールを用いて、モデルがこのプロセスを反復試行させるとルールの統合が起これ、内部で新しいルールが自動的に作られる。図4は学習後に記憶検索に関わるルールが統合された際のプロセスである。記憶の検索に関するルールが減り、大幅に思考時間が減る。つまり、人間的な慣れと同様の振る舞いをする [1]。そのため、結果的にモデルの思考から次の行動までの時間が減り反応が早くなる。

上記の仕組みを先述した「鬼ごっこ迷路ゲーム」に当てはめれば、モデルはゲーム初期において、環境地図の記憶、他者の過去の振る舞いの記憶などの検索を度々行う、ゲームが進行するに従って、それらの記憶検索が必要なくなる。その結果、モデルのゲームの遂行が効率化されることになる。

#### 4.1.2 ユーティリティ

ACT-R における効用 (utility) とは、競合解消（特定の状況において発火可能な複数のルールから1つを選択すること）に利用されるパラメータであり、個別のルールに付与される。ACT-R の「ユーティリティモジュール」は、効用値の更新を制御するモジュールであり、効用関数の計算に報酬を用いる。

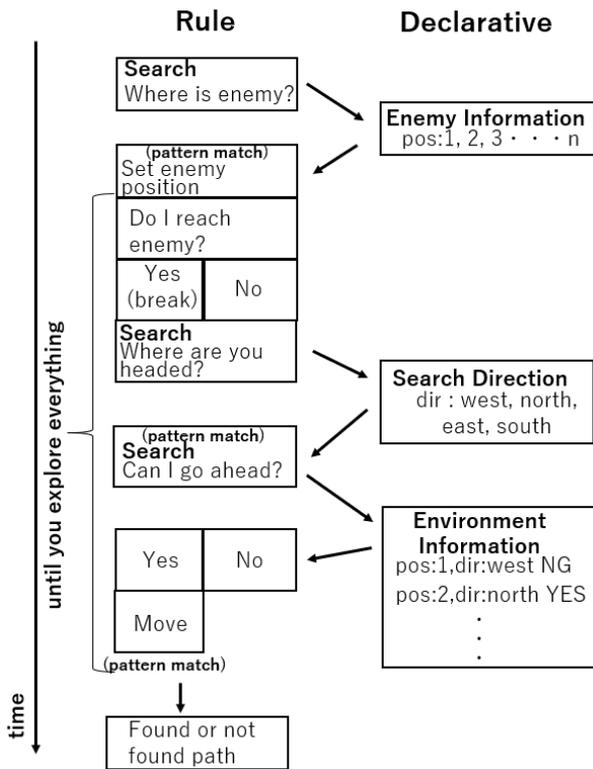


図3 学習前のプロセス

4.1.3 インタラクション継続の「楽しさ」のモデル

ACT-Rのこれらの機能を用いて、「楽しさ」と「飽き」を持つ認知モデルを提案する。図5は本ゲーム課題におけるインタラクション継続の提案モデルを示したものである。ゲーム内の各ラウンドの開始時において、ゲーム続行、ゲーム終了の判断（競合解消）を行う。ゲーム続行が選択された後、ゲーム内における種々の操作に関わるルール（地図の探索、他者の振る舞いの予測など）が発火した後にラウンドが終了し、新たなラウンドがスタートする。

上記のプロセスにおいて、ゲーム続行ルールの効用値の初期値は、ゲーム終了ルールの効用値の初期値より高く設定されると考える。ゲーム開始時に、人間はある程度はゲーム継続の意思があると考えられるためである。この初期状態からの「飽き」のプロセスは、各ラウンドの終了を認識するルールの発火を負の報酬のトリガーとすることでモデル化できる。ラウンド終了時に負の報酬が与えられることで、そのラウンドにおいて競合解消の結果として発火したルール（すなわちゲーム続行ルール）の効用値が低下し、ゲーム終了ルールが選択される確率が増加する。

「飽き」を抑制し、ゲームを継続させる条件を検討するためには、ゲームにおける「楽しさ」のモデル

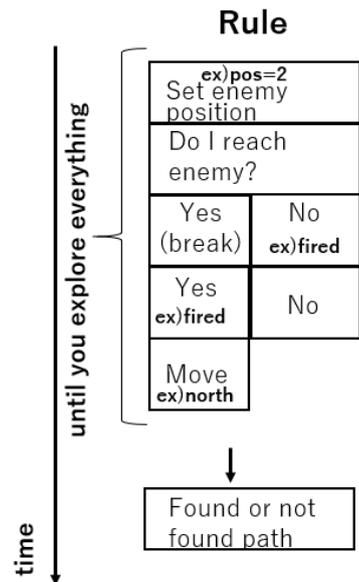


図4 学習後のプロセス

が必要である。ゲーム中に「楽しい」と感じるプロセスが生じた時に正の報酬がトリガリングされれば、ゲーム続行ルールの効用値は高い値を保ち、継続的なインタラクションが可能になる。本研究では、正の報酬のトリガーとなるルールを、ゲーム中の宣言的知識の検索成功（地図を思い出せた、他者の振る舞いの予測できたなど）に付随して発火するルールと定義する。宣言的知識の検索はルールのIF節（現在の状況）と宣言的知識内の記憶とのパターンマッチングであり、これに成功することはRaphaelによる楽しさの定義と整合する。ただし、このルールは繰り返し実行されると慣れ、つまりルールの統合が起こる。統合が起こるとルーティン化してしまい報酬を得られなくなり、ゲーム続行ルールの効用値が減少してゲーム終了ルールが発火するようになる。つまりこのモデルにおけるゲーム続行の要因は、パターンマッチの対象となる宣言的知識を見つけ続けられることとなる。

5. プロトタイプの実装とシミュレーション

5.1 概要

本研究の目的は、インタラクションの持続条件を明らかにするために、迷路環境において互いに鬼ごっこをする認知モデルを実装することである。この目的を達成する前段階として、モデルが他者の位置を予測しつつ、動的に変化する移動先のパスのプランニングを行う必要がある。このようなモデルを構築する第一歩として、静的なスタートからゴールまでの経路をプランニングするモデルを構築した。モデルの宣言的知識に

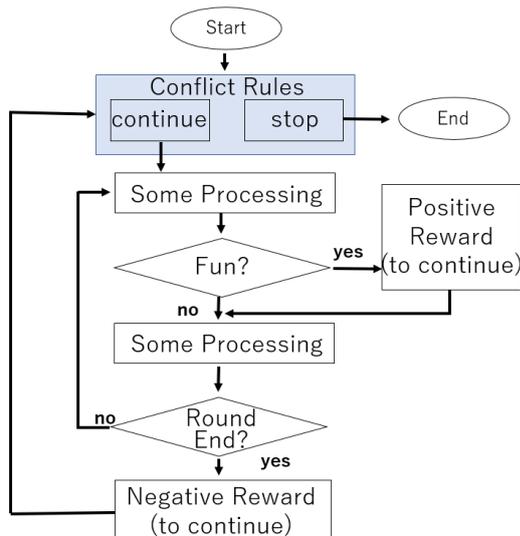


図5 タスク継続モデルのフローチャート

は、環境の地図が事前に記憶される。宣言的記憶内において、環境地図は曲がり角のつながりをグラフとして連結するトポロジカルマップとして表現される。モデルは記憶内のトポロジカルマップに対し、スタートからゴールまでのパスの探索を繰り返す。

この単純な迷路課題の中で、前節で示した楽しさや飽きのプロセスがどのように生じるのかを検討した。このモデルにおいて、パターンマッチに付随する「楽しさ」は、現在の状況からゴールまでの正しい道を思い出すことと定める（パスの発見時に発火するルールを正の報酬のトリガーとする）。また、パス探索の終了を課題の継続のモチベーションの減衰と定義する。つまり、ゴール探索の成功、失敗によらず、ラウンドの終了時に発火するルールを負の報酬のトリガーとする。ラウンド中に楽しさを表すルールが発火せずに、ラウンド終了に伴う負のトリガリングが生じ続けることで、継続ルールの効用値は減少していく。そして、継続ルールの効用値が、終了ルールの効用値を下回った時に、終了ルールが発火し、課題が終了することになる。なお、本モデルは、宣言的知識のトポロジカルマップに対するDFS(depth-first search[8])によって環境探索を行う。DFSはパスに関する知識を再帰的に検索するヒューリスティックであり、コンパイルによって探索プロセスが顕著に変化すると考える。以下、本モデルの実装に関わる詳細を示す。なお、実装の詳細は、シミュレーション結果の再現性を保証するために記載するものであり、5.5までスキップしても本論文の論旨を追うことは可能である。

## 5.2 迷路環境

図6はプロトタイプモデルに探索させる迷路環境である。3x3の空間内の各曲がり角に1から9まで番号を振り、モデルに空間内の場所を認識させる。本課題において、1番をスタート（エージェントの初期位置）、7番をゴール（非鬼役ならゴール、鬼役なら予想される非鬼役の位置）と静的に決める。また、モデルは、各曲がり角において西、北、東、南を0,1,2,3と認識する。この環境において、DFSで探索する場合、正解を見つけるモデルの移動は、1→2→3→2→5→6→9→6→5→8→7となる。この空間は、図2に示した人間の実験課題としては小さいものであるが、今回のプロトタイプモデルの目的は、「楽しさ」と「飽き」を要因としたタスク継続のモデル化であり、課題の環境そのものは、シンプルなものとした。

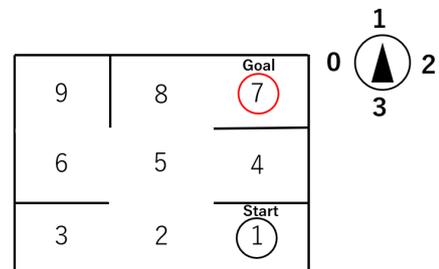


図6 迷路環境

## 5.3 モジュール

2.2節で前述した通りACT-Rは複数の独立したモジュールから構成されている。本モデルを実装するため、これらモジュールの中からゴールモジュール、宣言的モジュール、プロダクションモジュール、イマジナルモジュールを利用した。以下、これらのモジュールの概説と本研究のモデルにおける役割を示す。

### 5.3.1 ゴールモジュール

本モジュールは、モデルの現状態を保持するモジュールである。このモジュールは、対応するバッファを1つ有しており、そのバッファにチャンクと呼ばれる知識のかたまりを対応させることができる。この時、そのバッファはモデルの現状態となる。モデルは状態を参照、更新し動作を行う。

本モデルにおいて、チャンクは複数のスロットをもつ。本モデルにおけるスロットには、初期状態、現在位置、ゴールの座標が含まれる。また、各座標における向き、後ろ向き、スタックのトップのチャンク名などが含まれる。

### 5.3.2 宣言的モジュール

本モジュールは、人間の宣言的知識に対応するモジュールである。このモジュールは、対応するバッファを1つ有しており、宣言的知識から記憶を想起した際に、想起したチャンクをこのバッファに格納する。この時モデルは、このバッファを参照し、モデルの現状態の一部とすることができる。

本モデルにおいて、図6の環境についての宣言的知識を、モデルは事前に保持していた。各セルの探索順序の知識は、現在の方向と、次に向く方向で構成されている。まず西から探索し、北、東、南方向の探索を行う。南方向の探索が行われるとそのセルの探索が終了となる。図7は探索順序を実現するための知識の繋がりを表したものである。これら知識の繋がりによって、各時点のモデルは次に探索すべき方向を検索することができる。

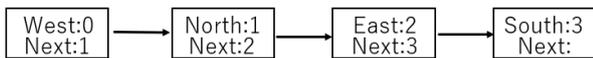


図7 探索の順番

また、迷路環境のトポロジカルマップの知識は、曲がり角番号、次の曲がり角番号、向き、後ろ向きで構成される。曲がり角番号とモデルの向きを用いることで、次の曲がり角番号がわかる。後ろ向きの情報は、曲がり角探索中に途中で後戻りしないようにするためのものである。図8は、迷路環境のトポロジカルマップを実現するために知識の繋がりを表したものである。これらの繋がりによって、迷路環境のモデル化を行った。

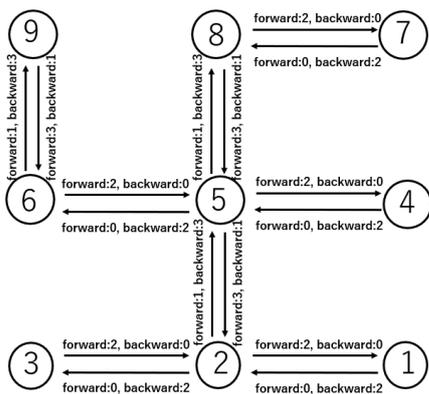


図8 環境のトポロジカルマップ

### 5.3.3 イマジナルモジュール

本モジュールは、新しいチャンクを生成し、宣言的知識に格納するためのモジュールである。このモ

ジュールは、対応するバッファを1つ有しており、その格納されたチャンクが宣言的モジュールに記憶される。イマジナルモジュールにおけるバッファに、他のモジュールが保持するチャンクを格納することで、複数のチャンクを組み合わせた新たなチャンクを構成することができる。例えばゴールに格納される現在位置と宣言的モジュールから検索された過去の記憶に関するチャンクを対応づけることができる。チャンク生成時に、時間的コストがかかるため、人がチャンクを生成する際の動作を模倣する事ができる。

本モデルにおいて、DFSで用いるスタック構造に関するチャンクの生成に用いている。詳細は後述5.4.1節に記述する。

### 5.3.4 ルールモジュール

本モジュールは、手続き的知識に対応するモジュールである。モデルの状態によって発火する条件(IF)と、発火後にモデルの状態を変更する(THEN)ルールを記述する。

本モデルは課題を開始する前に継続するか止めるかの判断を行う。継続ルールが発火すれば課題を継続し、終了ルールが発火すれば課題は終了となる。モデルは課題継続後、初期位置からゴール位置までの探索を行う。各位置において、モデルは西、北、東、南の順番で探索を行う。本課題ではDFSを採用しているため、現在地とモデルの向きから宣言的知識に対して、その先の曲がり角の情報を思い出せるかどうかの問い合わせを行う。問い合わせた結果、思い出せたら、モデルは現在の状況をスタックに積み、場所を移動し、その場所の左方向から探索を開始する。これを再帰的に繰り返す。もし、ゴール位置を発見した場合は探索成功となる。ある場所のすべての方向の探索が完了したら、スタックから1つ前の状態をポップし、その状態を用いて1つ前の場所に戻り各方向の続きを行う。もしスタックが空の状態でもポップした場合は探索失敗として課題終了となる。

## 5.4 DFSの実装

DFSはバックトラックを用いてグラフを探索するアルゴリズムである。グラフの探索初期位置から探索できるところまで探索を行い、目的地がなければ1つ前に戻り、他の方向を探索する(バックトラック)。本課題の開始の際に、モデルの宣言的知識に迷路環境をトポロジカルマップとして記述している。そのため、モデルは、それをグラフとして見立て迷路探索を行う。

上記の通り DFS の実装には、バックトラック機能が必要である。通常、手続き型などのプログラム言語においては、関数を再帰的に呼び出し探索を行うか、また探索の繰り返し処理の中に、バックトラック後の状態の復元のために現状態をスタックに積みながら探索を行うよう実装する。

ACT-R のモデルの実装において関数のような機能は存在しない。ACT-R のモデルは、基本的に、状態を変化させ、その状態によって発火するルールとチャンク生成、記憶、想起をしながら動作する。したがってバックトラック機能の実装のためには、これら機能を用いた、チャンクによるスタック構造と現状態をスタックに退避、復元する機能が必要である。

#### 5.4.1 チャンクを用いたスタック構造

図 9 は、本モデルにおいて実装したチャンクを用いたスタック機能である。本モデルでは、CHUNK-X をスタックに積んでいる。CHUNK-X の ARG1 にスタックの前のチャンク名が格納され、ARG2 に CHUNK-S-X のチャンク名が格納されている。CHUNK-S-X は、モデルの状態、つまり、曲がり角の番号、モデルの前の向きの方角、モデルの後ろの向きの方角が格納されている。また、モデルは、自身の状態の 1 つとしてスタックの最上部 (tail) のチャンク名を持つ。これらの構造と、前述したモジュールを用いてチャンクの生成、記憶、想起を行いスタック構造を実現している。

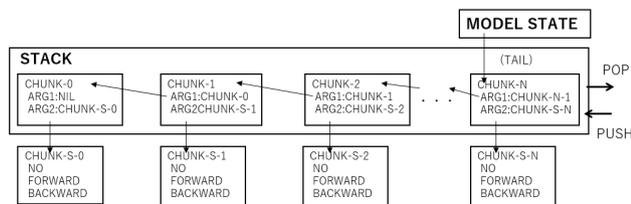


図 9 本モデルのチャンクを用いたスタック構造

スタックにおけるプッシュ機能は、モデルの状態の更新、チャンクの生成と記憶を用いて実現している。スタックに積むチャンクを生成する際に、モデルが持つスタックの最上部のチャンク名を 1 つ前のチャンク名とする。加えてモデルは、最上部のチャンク名をこのチャンク名に更新することでスタックにおけるプッシュ機能を実現している。なお、これらの生成後のチャンクは宣言的知識に記憶されるため後ほど想起することが可能である。

スタックにおけるポップ機能は、モデルの状態の更新とチャンクの想起によって実現している。モデルはスタックの最上部のチャンク名をもとにチャンクの内

部を思い出す。内部にはスタックの最上部の 1 つ前のチャンク名があるので、このチャンク名をモデルが持つスタックの最上部のチャンク名として更新することでポップ機能を実現している。

#### 5.4.2 現状態の退避と復元

現状態の退避は、モデルが向いている方向に存在する曲がり角に移動する前に状態をスタックにプッシュする事で実現している。これによって、移動して行く際の状態がスタックに保持される。

状態の復元は、任意の曲がり角の探索終了 (西, 北, 東, 南) した際に、スタックから状態をポップし、モデルの状態を更新することで実現している。これによって、移動前の状態に戻り、移動前の次の方向から探索を開始することができる。

### 5.5 シミュレーション

図 10 はモデルのシミュレーションの結果である。継続ルールの効用値の初期値を 10, 終了ルールの効用値の初期値を 5 とした。その他パラメータは ans (activation noise level) = 0.5, blt (base-level learning) = 0.5, blc (base-level constant) = 10, rt (retrieval threshold) = -0.5, egs (expected gain s) = 10 を設定した。またパスの発見 (DFS の成功) によってトリガーされる報酬値を 1 から 20 まで変化させ、各報酬値に対して 100 回、課題の継続シミュレーションを行った。報酬値を変化させた際に課題継続数の推移をグラフ化したものである。報酬値が大きいくほど継続数が増えていることを確認できる。

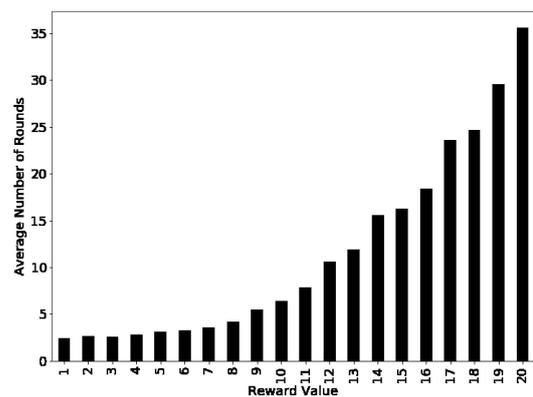


図 10 100 回の試行のうちの継続平均回数

図 11 は 100 回の実行結果から、20 回分を無作為に抽出し、課題実行回数に対して課題の実行時間の推移をグラフ化したものである。グラフ中で色付けされた各線が各試行の結果を示し、黒のダッシュ線は、平均

の推移を表す。ラウンド数が大きくなるにつれ課題実行時間が小さくなっていることが確認できる。このことは、コンパイルによって宣言的知識が検索される頻度が減少し、タスク遂行にかかる時間が減少したと考える。汎用的なアルゴリズムである再帰的なバックトラックがコンパイルされることで、今回の課題環境に特化した行動が学習されたと考えることができる。また、報酬値の低い条件においては、十分に課題時間が減少しない段階でモデルの実行がストップしていることも確認できる。

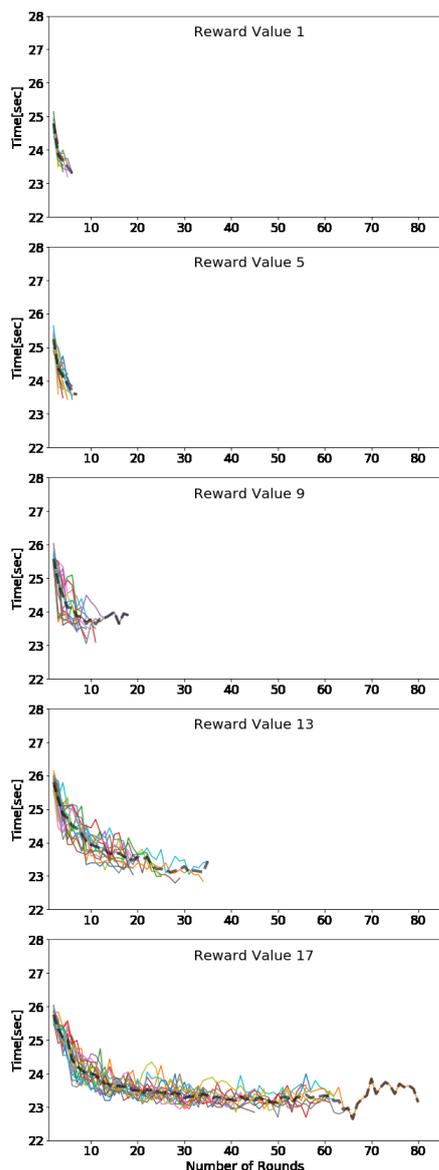


図 11 各報酬値に対して課題完了時間の遷移

図 12 は各報酬値条件における効用値の変動を、図 12 と同様のフォーマットで示している。グラフより、報酬値が低い条件では線形に効用値が減少することがわかる。それに対して、報酬が一定の値以上（10 を超

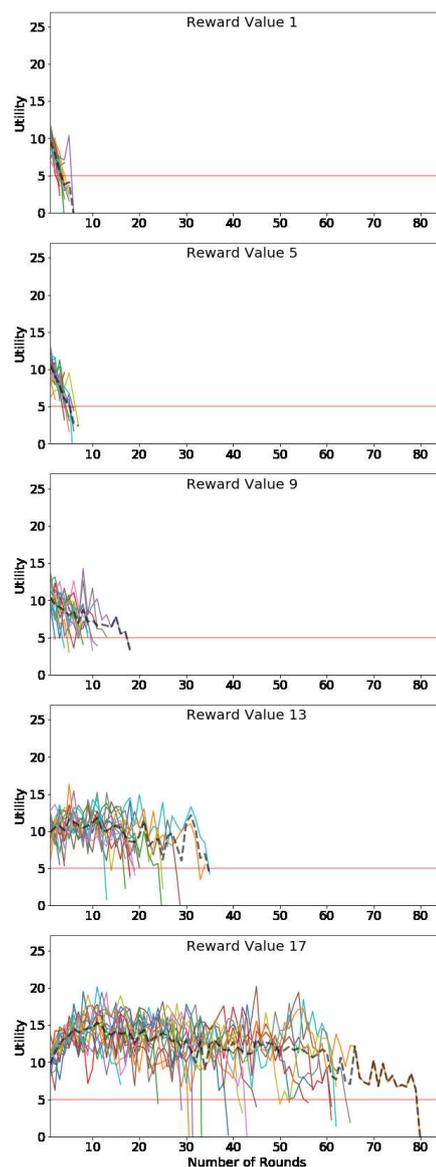


図 12 各報酬値に対して継続ルールの効用値の遷移 (赤の横線は終了ルールの値)

える値) の条件においては、課題の遂行に応じて、効用値の向上が散発的に生じ、ラウンドが継続することがわかる。すなわち、課題の遂行において、新たなパターンの発見が生じた際には、報酬を受け取り継続数を伸ばしている事が確認できる。

## 6. まとめ

本研究の目的は、ACT-R から提供されるプリミティブな認知プロセスの集積によって、インタラクションの持続がどのように成し遂げられるのかを明らかにすることであった。この目的の達成のために、実際に人間とエージェントが仮想環境においてインタラクションする課題を想定し、実験において得られた人間の行

動特性を再現するモデルを議論した。鬼ごっこ迷路課題を利用した実験の結果、勝敗がタスク継続に影響しないこと、対戦相手が自分とは異なる個人特性を有することがインタラクションの継続に寄与することが明らかになった。

本研究で提案するモデルは、課題中にモデルが感じる「楽しさ」を、状況と宣言的記憶とのパターンマッチングの成功によってモデル化するものである。課題環境に関わる先読みや他者の反応の予測には、宣言的知識が利用されると仮定する。よって、これらのプロセスが成功することによって、モデルは課題を継続させる。そして、この利用がコンパイルによってスキップされることで、モデルは課題に対する「飽き」を生じさせ、インタラクション課題を終了する。

このようなモデル化は、従来の IMRL において示されてきた予測可能性の議論をよりプリミティブなレベルで説明するものである。また、ACT-R による過去の研究においてもパターンマッチによる報酬の付与や、コンパイルによるタスクへの飽きを表現するものはなく、本研究において示したモデルは、オリジナルのものといえる。

提案されたアルゴリズムは、静的な状況における迷路探索の課題に適用された。そのための前段階としてプロトタイプモデルを作成し、シミュレーションを行いモデルの振る舞いを確認した。得られたシミュレーション結果から、パターンマッチを伴うルールに対する報酬の付与が、タスク継続のルールの効用値を上昇させ、長期間にわたる課題遂行を持続させることが示された。この結果から、本研究において提案するパターンマッチによる楽しさのモデル化が、インタラクション継続の必要条件ということは言えそうである。

今後の課題として、本研究におけるモデルをモデル同士がインタラクションできるように拡張する必要がある。しかし、現在モデル同士がインタラクションを行うインターフェイスを持たない。この過程においてインタラクションをするためのインターフェイスやデータの保管方法を検討する必要があるだろう。

## 7. 参考文献

### 文献

- [1] J. R. Anderson. Skill acquisition: Compilation of weak-method problem situations. *Psychological review*, Vol. 94, No. 2, p. 192, 1987.
- [2] J. R. Anderson. *How Can the Human Mind Occur in the Physical Universe*. OxfordPress, 2007.
- [3] J.R. Anderson. Knowledge compilation: The general learning mechanism. *Machine learning: An artificial intelligence approach*, Vol. 2, pp. 289–310, 1986.

- [4] S. Baron-Cohen. *The essential difference: The Truth About The Male And Female Brain*. Basic books, 2003.
- [5] T. W. Bickmore and R. W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, Vol. 12, No. 2, pp. 293–327, June 2005.
- [6] D. Bothell. Act-r 7.6+ reference manual. 2004.
- [7] N. Chentanez, A. G. Barto, and S. P. Singh. Intrinsically motivated reinforcement learning. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*, pp. 1281–1288. MITPress, 2005.
- [8] T.H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 3rd. edition, 2009.
- [9] J.J. Gibson. *The ecological approach to visual perception: classic edition*. PsychologyPress, 2014.
- [10] H. Komeda, H. Kosaka, DN. Saito, Y. Mano, M. Jung, T. Fujii, HT. Yanaka, T. Munesue, M. Ishitobi, M.Sato, and H. Okazawa. Autistic empathy toward autistic others. *Social Cognitive and Affective Neuroscience*, Vol. 10, No. 2, pp. 145–152, 10 2014.
- [11] R. Koster. *Theory of Fun for Game Design*. ParaglyphPr, 12 2004.
- [12] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves and I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
- [13] A. Newell. Production systems: Models of control structures. In *Visual information processing*, pp. 463–526. Elsevier, 1973.
- [14] T. Nozawa and T. Kondo. Autonomous adaptive agent with intrinsic motivation for sustainable hai\*. *JILSA*, Vol. 2, pp. 167–178, 01 2010.
- [15] A. H. Qureshi, Y. Nakamura, Y. Yoshikawa, and H. Ishiguro. Intrinsically motivated reinforcement learning for humanrobot interaction in the real-world. *Neural Networks*, Vol. 107, pp. 23 – 33, 2018. Special issue on deep reinforcement learning.
- [16] D. Reitter and C. Lebiere. A cognitive model of spatial path-planning. *Computational and Mathematical Organization Theory*, Vol. 16, No. 3, pp. 220–245, 2010.
- [17] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MITPress, 2 1998.
- [18] M. van de Velde, M. van Vugt, and N. Taatgen. Modelling the effect of depression on working memory. In *Proceedings of the 17th Annual Meeting of International Conference on Cognitive Modeling*, pp. 199–204, 2018.
- [19] F. Wai-Tat and J. R. Anderson. From recurrent choice to skill learning: A reinforcement-learning model. *Journal of experimental psychology. General*, Vol. 135, pp. 184–206, 6 2006.
- [20] C. J. C. H. Watkins. Learning from delayed rewards. 1989.
- [21] 岡真奈美, 森田純哉, 大本義正. インタラクションを持続させる個人特性—システム化と共感に注目した検討—. 2018.