

エージェントによる内部状態推定の複雑度変化

Changes in Complexity of Internal State Inference Depending on Agency

栗野嘉隆^{1*} 山田誠二^{2,3,1} 寺田和憲⁴
Yoshitaka Awano¹ Seiji Yamada^{2,3,1} Kazunori Terada⁴

¹ 東京工業大学 情報理工学院 情報工学系 知能情報コース

¹ Tokyo Institute of Technology

² 国立情報学研究所

² National Institute of Informatics

³ 総合研究大学院大学

³ the Graduate University for Advanced Studies(SOKENDAI)

⁴ 岐阜大学

⁴ Gifu University

Abstract: 人間の認知過程をモデル化するには、記憶の誤りや先入観などの認知バイアスに関する知見が重要である。先行研究において、人間とエージェントとのインタラクションにおける人間の認知バイアスは、エージェントの擬人的特性に依存することが示唆されている。そのため、人間とエージェントとのインタラクションにおける人間の認知過程をモデル化するには、エージェントの擬人的特性と人間の認知バイアスの関係性を考慮する必要があると考えられる。一方、認知科学において、人間は実世界との類似度が高い問題を解いた場合の方が、正解率が高まることが知られている。実世界との類似度が高い問題に人間が取り組んだ際、問題を解く過程で認知バイアスが生じ、正解率が高まると考えられる。そこで、本研究では有限状態オートマトンの学習に焦点を当て、エージェント自身の擬人的特性の差異を含め、異なる状況下で人間が異なるエージェントとインタラクションを行ない、エージェントの内部状態遷移を有限状態オートマトンの形式で推定する際に、人間の推定パフォーマンスがどのように変化するかを実験的に解明することを目的とした。本実験では、独立変数として擬人的特性の異なる2種類のエージェントを設定し、それらのエージェントに基づき、実世界との類似度が異なる2種類のインタラクション場面を構築し、それらの差異が人間の推定パフォーマンスに与える影響、すなわち人間の認知バイアスとの関係性に注目した。実験結果より、実世界との類似度が高い問題の方が、参加者の推定パフォーマンスが優れていた。また、実世界の類似度が高い問題でのみ、参加者は有限状態オートマトンの内部状態の「意味」を見出し、規則性を理解しやすく、問題を解く上で難しさを感じていなかったと考えられる。これらのことから、有限状態オートマトンの学習の難易度は、インタラクションの内容の実世界との類似度にも影響され、実世界との類似度が高いおかげで内部状態の「意味」を見出すことができ、この「意味」の発見のプロセスが、認知バイアスによるものである可能性が示唆される。以上より、実世界との類似度が異なる内容の有限状態オートマトンを人間が学習する際、内容の違いに応じた認知バイアスが生じ、人間の推定パフォーマンスが変化するという仮説が実証された。

1 はじめに

人間の認知過程をモデル化する為には、記憶の誤りや先入観などの認知バイアス [1] に関する知見が重要で

ある。人間は日常生活の中で、あらゆる学習場面に遭遇する。それらの中にはミーリ・マシン [2] で表現できるものも多く、人間は認知機能を働かせることで学習を遂行する。その際に生じる認知バイアスに関する知見を得ることで、学習時の認知過程のモデル化が可能となる。また、例として認知モデルは、インタラクションデザインに応用される。例えば、階層メニューは認

*連絡先：東京工業大学 情報理工学院 情報工学系 知能情報コース

〒152-8550 東京都目黒区大岡山 2-12-1
E-mail: awano.y.ab@m.titech.ac.jp

知バイアスに基づいて階層深度の上限が設定される。

認知科学において、人間は実世界との類似度が高い問題を解いた場合の方が、正解率が高まることが知られている [3]。ここで、実世界の類似度が高いとは、問題の内容が日常的あるいは現実的であることを指す。実世界との類似度が高い問題に人間が取り組んだ場合、問題を解く過程で認知バイアスが生じることで、正解率が高まると考えられる。

一方、人間とエージェントとの競争状況下で人間がエージェントの行動戦略に適應する際には、エージェントの擬人的特性に応じた認知バイアスが生じることが示唆されている [4, 5]。[5]では人間とエージェントが行うシンプルな2人対戦ゲームにおいて、ゲーム途中にエージェントの行動戦略が変化する状況を対象とし、ゲーム途中ごとのプレイヤーの勝率を記録した。対戦相手として熊型ロボット、NAO(Aldebaran Robotics社製)、ラップトップPC、人間の4種類のエージェントを用いた。結果、人間エージェントと対戦した際、他の3種類のエージェントの場合よりも、エージェントの行動戦略が変化した後のプレイヤーの勝率が低下した。これより、人間とエージェントとのインタラクションにおける人間の認知過程をモデル化する上では、認知対象であるエージェントの擬人的特性と人間の認知バイアスとの関係性を検討する必要があると考えられる。この関係性を解明することで、人間の学習効率を向上させるエージェントやインタラクションデザインに指針を与えることが期待できる。

これらの背景から、本研究では、エージェント自身の擬人的特性の差異を含め、異なる状況下で人間がエージェントとインタラクションを行ない、エージェントの内部状態遷移を推定する際に、人間の推定パフォーマンスがどのように変化するかに注目する。

2 ミーリ・マシンおよび認知バイアス

本章では、ミーリ・マシンおよび認知バイアスについて説明する。

2.1 ミーリ・マシン

2.1.1 ミーリ・マシンの定義

ミーリ・マシン [2] は、出力が入力と状態に依存した決定性有限オートマトンであり、 $M = \langle I, O, Q, q_0, \delta, \lambda \rangle$ で表される。ここで、 I は入力集合、 O は出力集合、 Q は状態集合、 q_0 は開始状態、 δ は遷移関数 $S \times I \rightarrow S$ 、 λ は出力関数 $S \times I \rightarrow O$ である。ミーリ・マシンの例を図1に示す。この例では、 $I = \{a, b\}$ 、 $O = \{A, B, C\}$ 、 $Q = \{q_0, q_1, q_2\}$ である。開始状態 q_0 から始まり、例

えば状態 q_0 で a を入力すると、 A を出力して状態 q_0 に遷移 (ループ) し、状態 q_0 で b を入力すると、 B を出力して状態 q_1 に遷移する。これらは δ および λ を用いて、 $\delta(q_0, a) = q_0$ 、 $\lambda(q_0, a) = A$ 、 $\delta(q_0, b) = q_1$ 、 $\lambda(q_0, b) = B$ と表される。また、遷移は決定論的に行われ、各状態は必ず各入力に対してただ1つの遷移を持つ。すなわち、遷移数は (状態数 \times 入力数) となる。図1の場合、遷移数は (状態数 $3 \times$ 入力数 2) = 6 となる。加えて、 O の要素は全て、いずれかの遷移時に必ず出力される。

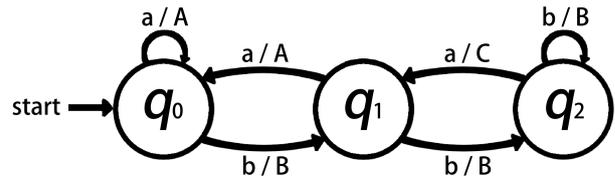


図1: ミーリ・マシンの例

2.1.2 ミーリ・マシン学習の手法

1987年、Angluin[6]は2種類のクエリを用いてミーリ・マシンを学習できる事を示した。この手法はMAT (Minimally Adequate Teacher) 学習と呼ばれる。教師はミーリ・マシン M を知っており、学習者は M の入力集合 I と出力集合 O のみ知っており、学習者は2種類のクエリを用いて M を学習する。1つはメンバーシップクエリ (Membership Query) と呼ばれ、学習者は入力シーケンス $\sigma \in I^+$ に対する出力を教師に尋ねる。教師は出力シーケンス $OUT(\sigma) \in O^+$ を返す。ここで、 $OUT: I^+ \rightarrow O^+$ である。もう1つは等価クエリ (Equivalence Query) と呼ばれ、学習者は仮説ミーリ・マシン H を作成し、 H と M が等価であるかを教師に尋ねる。等価な場合場合、教師は「yes」と答えて終了、そうでなければ教師は「no」と答え、 H と M を区別する反例 $z \in I^+$ を学習者に返す。これらのクエリを用いることで、学習者は M を学習できることが保証されている。

MAT学習に基づくミーリ・マシンの学習アルゴリズムがいくつか提案されており、Angluinの L^* アルゴリズム [6] や、IsbernerらのTTTアルゴリズム [7, 8] などがある。多くの学習アルゴリズムにおけるメンバーシップクエリの必要数は、入力数に比例して増加し、状態数に対して二次的に増加する [7]。仮説ミーリ・マシン H の状態数が n 、学習対象のミーリ・マシン M の状態数が n' で、入力数が k の時、最悪の場合 $k^{(n-n')}$ 個のメンバーシップクエリが必要である [9]。TTTアルゴリズムの場合、 M の状態数が n 個、入力数が k 、最長の反例の長さが m である時、最悪の場合、 $O(n)$

個の等価クエリと $O(kn^2 + n \log m)$ 個のメンバーシップクエリが必要である。

以上のことから、ミーリ・マシン学習の複雑度は、状態数や入力数に依存すると考えられる。

2.2 認知バイアス

人間は認知を行う際に、記憶の誤りや先入観などにより、認知や思考・推論が歪められてしまうことがある。これらの原因となるものを総称して認知バイアス [1] と呼び、人間の認知過程を正確にモデル化する上で、認知バイアスに関する理解は不可欠である。認知バイアスは人間が学習を行う際にも生じ、学習効率を上昇あるいは低下させる原因となる。

認知バイアスは様々な観点から分類される。その内の一種に確証バイアス [10] がある。これは、仮説や信念を検証する際、それを支持する情報ばかりを集め、反証する情報を無視あるいは集めようとしない傾向を指す。確証バイアスの例として、ウェイソン選択課題 [11] がある。例えば、図 2 に示す 4 枚のカードについて、以下の課題が与えられたとする。

**各カードは片面に数字が書かれ、
もう片面には色が塗られている。
「カードの片面に偶数が書かれているならば、
もう片面は赤い」
という仮説を確かめるためにひっくり返す必要が
あるカードはどれか？**

この回答として多いのは、「8 のカード」あるいは「8 のカードと赤色のカード」である。しかし、これらは合理的ではない。何故なら、与えられた仮説の反例になり得るのは「片面に偶数が書かれていて、かつ裏面が赤色でないカード」すなわち「8 のカードと茶色のカード」だけであり、その他の組合せは仮説の検証には全く役立たない。多くの人がこのような問題に誤答することは、確証バイアスの結果として説明される。

こうした認知バイアスについて体系的に解明されれば、人間がどのように失敗をするかを予測し、失敗を回避する方策を立てることができる。

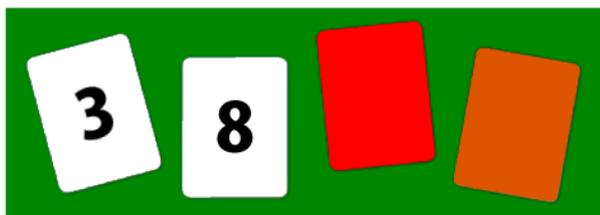


図 2: ウェイソン選択課題

3 人間の学習プロセスのモデル化

本研究では、人間とエージェントのインタラクションとして、人間がエージェントに入力を与え、それに対してエージェントが出力を返す形式を採用する。その中で、いくつかの出力の内容が分からなくなっているので、その出力の内容を人間が推定し、エージェントの正しい内部状態遷移をミーリ・マシンとして回答する。

一部の出力が未知のミーリ・マシンの例を図 3 に示す。これは図 1 と同じミーリ・マシンだが、6 つの出力のうち 3 つ、すなわち $\lambda(q_0, b)$, $\lambda(q_1, a)$, $\lambda(q_2, a)$ が分からなくなっている (図 3 中の () の部分)。このようなミーリ・マシンに対し、未知の出力の内容を人間が推定し、本来の正しいミーリ・マシンを完成させる。

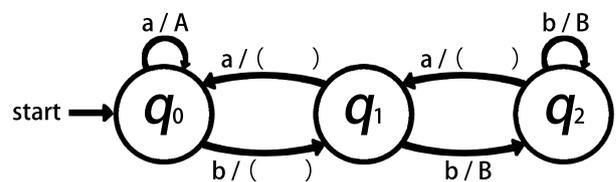


図 3: 一部の出力が未知のミーリ・マシンの例

3.1 学習プロセスのモデル

人間の学習プロセスは、学習が行われている状態と、学習が完了した状態とに分けられる。2 つの状態における、推測すべき問題に対する正解数を簡略化したグラフを図 4 に示す。通常、人間が他者とインタラクションをする場合、相手のとった行動から相手の振舞い、すなわち内部状態遷移のモデルを同定し、それに応じて自分自身の行動を決定する。このモデルはミーリ・マシンとして表すことができ、自分の行動は入力、それに対する相手の反応は出力としてミーリ・マシン上で表現される。人間は他者とのインタラクションの中で、このような相手の内部状態遷移のモデルを構築する。この為、相手のいくつかの出力が観測できない場合、その出力の内容を推測する必要があるため、初めは欠けている出力に対する正解数は少ない。しかし、思考を巡らせる中で相手の内部状態遷移パターンを学習し、それに応じて欠けている出力に対する正解数が次第に増していく。やがて相手の内部状態遷移パターンの学習が完了すると、人間は相手の出力パターンを完全に把握するため、欠けている出力に対する正解数が上限値に収束する。ここで、学習を開始してから学習が完了するまでの経過時間 τ (図 4) の逆数 $\frac{1}{\tau}$ を、人間の学習速度とする。

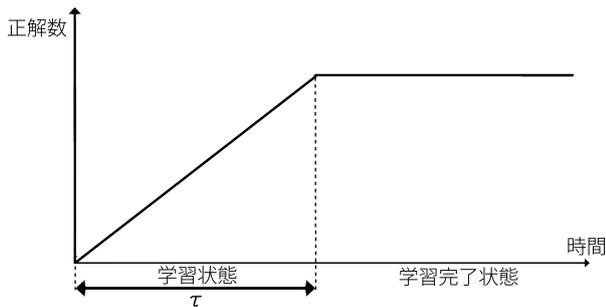


図 4: 学習状態および学習完了状態での出力の正解数

3.2 対象の違いによる学習速度の違い

先行研究 [4, 5] では、人間とエージェントとの競合ゲームにおけるエージェントの行動戦略に対する人間の適応速度と、エージェントの擬人的特性との関連性が実験的に検証され、人間がエージェントの行動戦略への適応を試みる際に、エージェントの擬人的特性に応じた認知バイアスが生じることが示唆されている。

これに基づき、本研究における人間とエージェントとのインタラクションにおいても、エージェントの擬人的特性やインタラクションの内容に応じた認知バイアスが生じ、人間の学習速度に変化が生じると仮定する。本研究は先行研究 [4, 5] と異なり、擬人的特性だけでなく、人間とのインタラクションの内容もエージェントごとに異なる。各インタラクションはそれぞれ、実世界との類似度に関する違いがあるものとする。

4 実験

本章では、本実験の目的とインタラクションの内容、実験条件および手順、評価方法、実験結果を述べる。

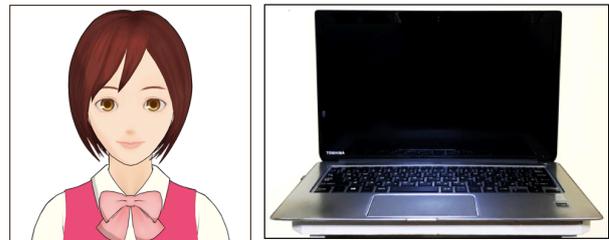
4.1 実験の目的

本実験では、擬人的特性の異なる 2 種類のエージェントを用意し、参加者は各エージェントと実世界との類似度が異なる内容のインタラクションを行い、エージェントの内部状態遷移をミーリ・マシンとして推定する。その際、エージェントの擬人的特性の差異を含む、インタラクションの内容の違いに応じた認知バイアスが生じ、人間の学習速度に変化が生じるという仮説を、本実験により検証する。

4.2 インタラクションの内容について

本実験では、擬人化エージェントおよびラップトップの 2 種類のエージェント (図 5) を使用する。擬人

化エージェントには、音声対話システムツールキット MMDAgent¹ の Web サイトで配布されている Mei を用いた。各インタラクションにおけるエージェントの内部状態遷移を、ミーリ・マシンとして出題する。なお、今回出題するミーリ・マシンは、人間が実験時間内に解ける難易度にする為に、入力数 2, 出力数 2, 状態数 4 に設定した。



(a) 擬人化エージェント

(b) ラップトップ

図 5: 実験で使用するエージェント

4.2.1 擬人化エージェントとのインタラクション

擬人化エージェントとのインタラクションについては、日常的なシチュエーションを構築した。具体的には、「学校内での教師と生徒のインタラクション場面」を想定し、実験参加者が教師役、擬人化エージェントが生徒役とする。擬人化エージェントとのインタラクション画面を図 6 に示す。ラップトップの画面上に教室の風景および擬人化エージェント、擬人化エージェントの現在の表情 (出力) および状態、擬人化エージェントへの入力のボタン、および開始状態 q_0 に戻るボタンが表示される。擬人化エージェントへの入力、および擬人化エージェントからの出力を図 7 に示す。入力は社会的行動 [12] から選択した。参加者は擬人化エージェントへの入力として、「キミはすごい」と言う、あるいは「もっと努力しろ」と言う、の 2 つのアクションを行うことができ、画面上の対応するボタンを押す (クリックする) ことで、入力が実行される。これに対し、擬人化エージェントは「無表情」「えがお」「なみだ目」「おびえ」のいずれかの表情を出力として返す。これらの表情は、基本感情 [13] に対応するものとして設定した。参加者からの入力が与えられると、擬人化エージェントは出力を返し、擬人化エージェントの内部状態遷移が起こる。

このインタラクションにおける擬人化エージェントの内部状態遷移を表すミーリ・マシンを図 8 に示す。このミーリ・マシン M_A は、入力集合 $I_A = \{\text{キミはすごい}, \text{もっと努力しろ}\}$, 出力集合 $O_A = \{\text{無表情}, \text{えがお}, \text{なみだ目}, \text{おびえ}\}$, 状態集合 $Q_A = \{q_0, q_1, q_2, q_3\}$, 開始

¹<http://www.mmdagent.jp/>

状態 q_0 , 遷移関数 $\delta_A: S_A \times I_A \rightarrow S_A$, 出力関数 $\lambda_A: S_A \times I_A \rightarrow O_A$ を用いて, $M_A = \langle I_A, O_A, Q_A, q_0, \delta_A, \lambda_A \rangle$ と表される. このうち, 出題用のミーリ・マシン (図 8(a)) では, 4つの出力 $\lambda_A(q_0, \text{キミはすごい})$, $\lambda_A(q_0, \text{もっと努力しろ})$, $\lambda_A(q_1, \text{キミはすごい})$, $\lambda_A(q_2, \text{もっと努力しろ})$ が分からなくなっている. これらの部分については, インタクション画面内で擬人化エージェントの姿が隠され, 出力に相当する擬人化エージェントの表情が分からなくなっている. 参加者はこれらの出力の内容を推定し, 正解のミーリ・マシン (図 8(b)) を当てる作業を行う. なお, 擬人化エージェントの状態は, 基本感情 (すなわち擬人化エージェントの各表情) に対応するものとして設定した. 例えば, 状態 q_1 は感情「喜び」に対応しており, q_1 に遷移する際の出力は全て「えがお」である. 最初, 擬人化エージェントは「無表情」から始まり, 連続で「キミはすごい」と入力すると段階的にポジティブな表情に遷移し (「おびえ」→「なみだ目」→「無表情」→「えがお」), 連続で「もっと努力しろ」と入力すると段階的にネガティブな表情に遷移する (「えがお」→「無表情」→「なみだ目」→「おびえ」).



図 6: 擬人化エージェントとのインタクション画面



図 7: 擬人化エージェントの入力および出力

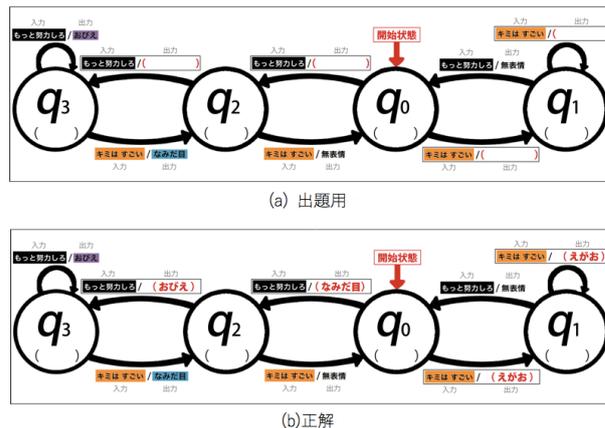


図 8: 擬人化エージェントの内部状態遷移を表すミーリ・マシン

4.2.2 ラップトップとのインタクション

ラップトップとのインタクションについては, 非日常的なシチュエーションを構築した. ラップトップとのインタクション画面を図 9 に示す. ラップトップの画面上に記号 (ラップトップの出力) およびラップトップの状態, ラップトップへの入力のボタン, および開始状態 q_0 に戻るボタンが表示される. ラップトップへの入力, およびラップトップからの出力を図 10 に示す. 参加者はラップトップへの入力として, 「はしをつかむ」あるいは「やさしくなでる」の2つのアクションを行うことができ, 画面上の対応するボタンを押すことで, 入力が実行される. これに対し, ラップトップは「 \wedge 」「 \int 」「 \oplus 」「 \exists 」のいずれかの記号を出力として返す. 参加者からの入力が与えられると, ラップトップは出力を返し, ラップトップの内部状態遷移が起こる.

このインタクションにおけるラップトップの内部状態遷移を表すミーリ・マシンを図 11 に示す. このミーリ・マシン M_L は, 入力集合 $I_L = \{ \text{はしをつかむ}, \text{やさしくなでる} \}$, 出力集合 $O_L = \{ \wedge, \int, \oplus, \exists \}$, 状態集合 $Q_L = \{ q_0, q_1, q_2, q_3 \}$, 開始状態 q_0 , 遷移関数 $\delta_L: S_L \times I_L \rightarrow S_L$, 出力関数 $\lambda_L: S_L \times I_L \rightarrow O_L$ を用いて, $M_L = \langle I_L, O_L, Q_L, q_0, \delta_L, \lambda_L \rangle$ と表される. このうち, 出題用のミーリ・マシン (図 11(a)) では, 4つの出力 $\lambda_L(q_0, \text{はしをつかむ})$, $\lambda_L(q_0, \text{やさしくなでる})$, $\lambda_L(q_1, \text{はしをつかむ})$, $\lambda_L(q_2, \text{やさしくなでる})$ が分からなくなっている. これらの部分については, インタクション画面内で出力記号が隠され, 内容が分からなくなっている. 参加者はこれらの出力の内容を推定し, 正解のミーリ・マシン (図 11(b)) を当てる作業を行う. なお, ラップトップの状態は, 各出力記号に対応するものとして設定した. 例えば, 状態 q_0 は記号「 \wedge 」に対応しており, q_0 に遷移する際の出力は全て「 \wedge 」である.



図 9: ラップトップとのインタラクション画面



図 10: ラップトップの入力および出力

4.2.3 各エージェントのミーリ・マシンの対応関係

本実験における擬人化エージェントおよびラップトップのミーリ・マシンはそれぞれ対応関係にあり、擬人化エージェントのミーリ・マシン (図 8) における入力および出力をそれぞれ表 1 の規則に従って変換すると、ラップトップのミーリ・マシン (図 11) と一致する。L* アルゴリズム [6] 等の学習アルゴリズムを用いてミーリ・マシンを学習する場合、入力および出力はアルファベット等に変換される (一般的に、入力は小文字のアルファベット、出力は大文字のアルファベットの変換される)。この為、学習アルゴリズムを用いて本実験における擬人化エージェントおよびラップトップのミーリ・マシンを学習する場合、この 2 つのミーリ・マシンはどちらも同じ 1 つのミーリ・マシンに対応する為、学習速度に差が生じない。これに対し、人間がこれらのミーリ・マシンを学習する場合には、インタラクションの内容の違いに応じて、学習速度に差が生じると仮定する。

本実験で出題されるミーリ・マシン (図 8(a) および図 11(a)) は、8 つの出力のうち 4 つが分からなくなっているが、これらの出力の内容は以下の根拠から推定することができる。まず、状態 q_0 に遷移する際の出力が全て同じ、すなわち $\lambda_A(q_1, \text{もっと努力しろ}) = \lambda_A(q_2, \text{キミはすごい}) = \lambda_L(q_1, \text{えがお}) = \lambda_A(q_2, \text{はしをつかむ}) = \int$ であり、このことから 1 つの状態は 1 つの出力に対応することが推定できる。また、明らかになっている 4 つの出力のうち、 q_0, q_2, q_3 に遷移する

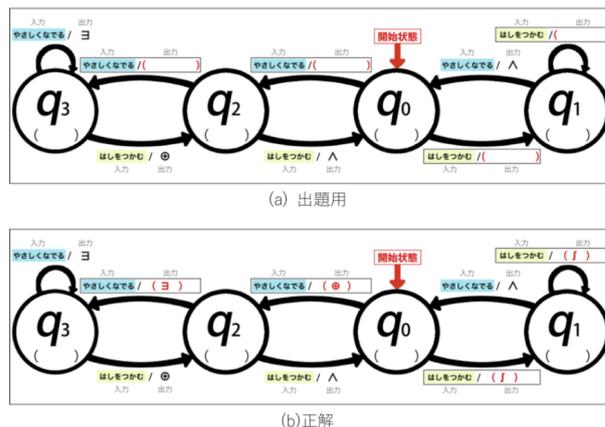


図 11: ラップトップの内部状態遷移を表すミーリ・マシン

際の出力が 3 つとも全て異なっており、このことから q_1 に遷移する際にはそれら 3 つとは異なる出力が返されることが推測できる。これらの根拠を元に、正解のミーリ・マシン (図 8(b) および図 11(b)) を推定することができる。これに加えて、擬人化エージェント条件ではインタラクションの内容が実世界との類似度が高いため、学習時に認知バイアスが生じ、ラップトップ条件に比べて学習速度が速くなると仮定する。

表 1: 擬人化エージェントとラップトップの入力および出力の対応関係

	擬人化エージェント	ラップトップ
入力	キミはすごい	はしをつかむ
	もっと努力しろ	やさしくなでる
出力	無表情	∧
	えがお	∫
	なみだ目	⊕
	おびえ	≡

4.3 実験条件・手順

本実験では、参加者はまず例題を解き、その後で擬人化エージェント条件およびラップトップ条件の問題

を解いた。どちらの問題を先に出題するかは参加者ごとに分け、先に擬人化エージェント条件の問題を解くものをA条件、先にラップトップ条件の問題を解くものをL条件と名付けた。今回は2要因混合計画の実験であり、要因は順序要因（水準：A条件およびL条件）および擬人化エージェント/ラップトップ要因（水準：擬人化エージェント条件およびラップトップ条件）である。これらの要因が独立変数であり、従属変数は各エージェントのミーリ・マシンに対する、学習速度を含む参加者の推定パフォーマンスである。

本実験は、24歳から55歳までの30名の参加者（男性16名、女性14名）で実施した。参加者の平均年齢（±標準偏差）は42.57(±9.23)歳であり、A条件は15名（男性8名、女性7名、平均年齢（±標準偏差）41.93(±9.59)）、L条件は15名（男性8名、女性7名、平均年齢（±標準偏差）43.20(±8.80)）で実験を行なった。参加者には実験の報酬として、1,500円相当のQUOカードを提供した。なお、実験では各エージェントおよび内部状態遷移について、参加者により馴染みのある名前として、擬人化エージェントを「エージェント」、ラップトップを「ノートパソコン」、内部状態遷移を「状態変化」と呼称した。

例題の内容はラップトップとのインタラクションで、実世界との類似度が高い問題として設定した。ラップトップとのインタラクション画面を図12に示す。画面構成はラップトップ条件のもの（図9）と同じである。ラップトップへの入力、およびラップトップからの出力を図13に示す。参加者はラップトップへの入力として、「半回転」あるいは「一回転」の2つのアクションを行うことができ、画面上の対応するボタンを押すことで、入力が実行される。これに対し、ラップトップは「↑」「↓」のいずれかの記号（矢印）を出力として返す。参加者からの入力が与えられると、ラップトップは出力を返し、ラップトップの内部状態遷移が起こる。このインタラクションにおけるラップトップの内部状態遷移を表すミーリ・マシンを図14に示す。このミーリ・マシン M_E は、入力集合 $I_E = \{ \text{半回転}, \text{一回転} \}$ 、出力集合 $O_E = \{ \uparrow, \downarrow \}$ 、状態集合 $Q_E = \{ q_0, q_1 \}$ 、開始状態 q_0 、遷移関数 $\delta_E: S_E \times I_E \rightarrow S_E$ 、出力関数 $\lambda_E: S_E \times I_E \rightarrow O_E$ を用いて、 $M_E = \langle I_E, O_E, Q_E, q_0, \delta_E, \lambda_E \rangle$ と表される。このうち、出題用のミーリ・マシン（図14(a)）では、2つの出力 $\lambda_E(q_1, \text{半回転})$ 、 $\lambda_E(q_1, \text{一回転})$ が分からなくなっている。これらの部分については、インタラクション画面内で出力記号が隠され、内容が分からなくなっている。参加者はこれらの出力の内容を推定し、正解のミーリ・マシン（図14(b)）を当てる作業を行う。なお、ラップトップの状態は、各出力記号に対応するものとして設定した。例えば、状態 q_0 は記号「↑」に対応しており、 q_0 に遷移する際の出力は全て「↑」である。この問題では、入力「半回転」

あるいは「一回転」に応じて、出力される矢印の向きが決まる。例えば、「↑」が出力されている状態で「一回転」を入力すると再び「↑」が入力され、「半回転」を入力すると「↓」が出力される。この規則に基づき、未知の出力の内容を推定することができる。この規則は実世界における方角の規則に即しているため、この問題は実世界との類似度が高いと言える。

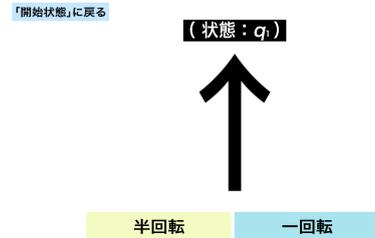


図 12: 例題におけるラップトップとのインタラクション画面

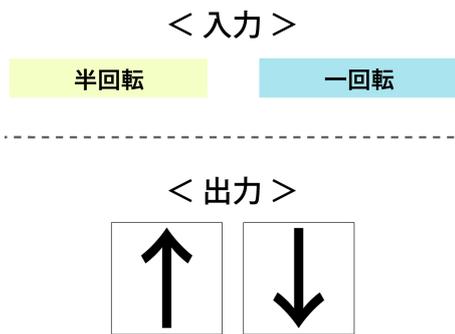


図 13: 例題におけるラップトップの入力および出力

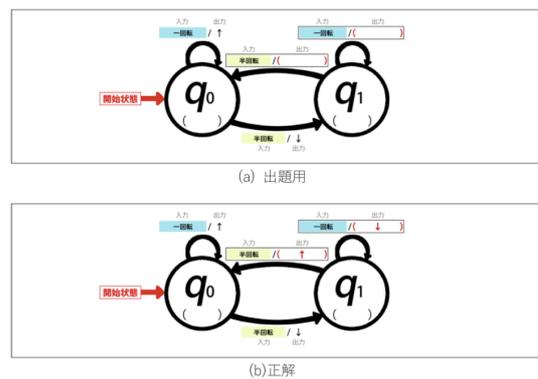


図 14: 例題におけるラップトップの内部状態遷移を表すミーリ・マシン

本実験におけるミーリ・マシン推定の手順について説明する。最初に、エージェントの内部状態遷移を表すミーリ・マシン（図8(a)あるいは図11(a)、図14(a)）がプリントされた解答用紙と、エージェントへの入力

およびエージェントからの出力（図7あるいは図10、図13）がプリントされた紙を参加者に渡す。参加者はこれらを受け取った後、実際にエージェントとインタラクションをし、解答用紙中のミーリ・マシンの通りにエージェントが動作することを確認する。この作業は、解答用紙中のミーリ・マシンが示すインタラクションのイメージを参加者に持たせる目的がある。今回の実験では、エージェントとのインタラクション時間を3分に設定した。その後、参加者は欠けている出力の内容を推定し、解答用紙中の赤い()にいずれかの出力を記入する。この際、ミーリ・マシンの各状態について、必要であれば状態内の黒い()に任意の言葉を記述できることを参加者に指示した。解答ができた場合、実験者が正解判定を行い、正解の場合はその時点で終了し、不正解の場合は解答の修正を参加者に指示する。この際、どの出力が誤っているか、あるいは出力がいくつ誤っているかは参加者に教えなかった。正解するまで、あるいは制限時間に達するまで、解答の作成および正解判定は何度でも行うことができる。今回の実験では、解答の制限時間を10分に設定した。本実験におけるミーリ・マシン推定の枠組みは、MAT学習[6]と同様であるが、本実験ではMAT学習における等価クエリ、すなわち正解判定に対する反例に相当するものが無い。また、MAT学習では全てのメンバーシップクエリ、すなわち入力に対する出力が必ず与えられるが、本実験では入力に対する一部の出力の内容は分からない。加えて、MAT学習ではメンバーシップクエリおよび等価クエリはそれぞれ任意のタイミングで行えるが、本実験ではエージェントへの入力は前半3分間のみ、解答の作成および正解判定は後半10分間のみ行うことができる。

4.4 評価方法

4.4.1 参加者のパフォーマンス

本実験では、各問題に対する参加者の正解率、参加者が正解のミーリ・マシンを学習するまでの経過時間（すなわち学習速度）、および各問題についての参加者の主観難易度のデータを収集し、これらを参加者のパフォーマンスとする。主観難易度については、問題を解くごとにその問題の難易度について、1~10の10段階で回答してもらった（1：簡単，10：難しい）。

4.4.2 アンケート

各エージェント条件の問題が全て終わった後、参加者にアンケートへの回答をもらった。

まず、本実験に対する参加者の主観的ワークロードを評価するため、NASA-TLXのワークロード概念に基

づく評価チェックリストである主観的メンタルワークロードチェックリスト²のうち、本実験と関連する質問に回答をもらった。各質問内容を表2に示す。質問は15項目あり、各項目に対する評価値は、「やさしかった」のようなポジティブな回答の場合は正数、「難しかった」のようなネガティブな回答の場合は負数とした。例えば1番上の項目への回答が「難しかった」に「とても当てはまる」の場合、評価値は-3となる。これらをもとに参加者の精神負担、身体負担、時間圧力、作業成績、フラストレーションを算出した。精神負担は「やさしかった」「単純だった」「頭を使う必要はなかった」「大ざっぱでよかった」への評価値の和、身体負担は「機敏な動作は必要ではなかった」への評価値、時間圧力は「時間に追われる感じはしなかった」への評価値、作業成績は「作業結果に満足である」「予想以上の結果が得られた」への評価値の和、フラストレーションは「不安を感じなかった」「いらいらしなかった」「楽しかった」「集中していた」への評価値の和である。今回はこれら5つの評価値の符号を逆にし、値が大きいかほどネガティブとした。

また、以下の質問について、参加者に回答をもらった。

Q1. エージェントの問題を解いた際、ノートパソコンの問題に比べてやる気ができましたか？

0. はい 1. いいえ 2. どちらでもない

Q2. ノートパソコンの問題において、画面上にエージェントが表示されている方がやる気がでると思いますか？

0. はい 1. いいえ 2. どちらでもない

Q3. エージェントおよびノートパソコンについて、どちらの問題の方がより日常的だと感じましたか？

0. エージェント 1. ノートパソコン 2. どちらでもない

Q4. 今回の問題中の「状態変化」について、何らかの規則性があると感じましたか？

0. はい 1. いいえ

（「0. はい」の場合、どのような規則性があると感じたかをご記入ください。）

Q5. エージェントの問題において、「状態」は何を表すものだと思いますか？

（自由記述）

Q6. ノートパソコンの問題において、「状態」は何を表すものだと思いますか？

（自由記述）

²<http://acpsy.hus.osaka-u.ac.jp/questionnaire.html>

Q7. エージェントの問題について、どのような点がやり易かった / 難しかったですか？
(自由記述)

Q8. ノートパソコンの問題について、どのような点がやり易かった / 難しかったですか？
(自由記述)

表 2: 主観的ワークロード評価に関する質問内容

	とてもあてはまる	かなりあてはまる	ややあてはまる	どちらでもない	ややあてはまらない	かなりあてはまらない	とてもあてはまらない
やさしかった	3	2	1	0	1	2	3
単純だった	3	2	1	0	1	2	3
頭を使う必要はなかった	3	2	1	0	1	2	3
大ざっぱでよかった	3	2	1	0	1	2	3
機敏な動作は必要ではなかった	3	2	1	0	1	2	3
時間に追われる感じはしなかった	3	2	1	0	1	2	3
作業結果に満足である	3	2	1	0	1	2	3
予想以上の結果が得られた	3	2	1	0	1	2	3
不安を感じなかった	3	2	1	0	1	2	3
いらいらしなかった	3	2	1	0	1	2	3
楽しかった	3	2	1	0	1	2	3
集中していた	3	2	1	0	1	2	3
得意である	3	2	1	0	1	2	3
好きである	3	2	1	0	1	2	3
慣れている	3	2	1	0	1	2	3
難しかった							
複雑だった							
頭を使う必要があった							
正確さが要求された							
機敏な動作が必要だった							
時間に追われる感じがした							
作業結果に不満である							
予想を下回る結果だった							
不安を感じた							
いらいらした							
つまらなかった							
気が散った							
苦手である							
嫌いである							
慣れていない							

4.5 実験結果

4.5.1 参加者のパフォーマンス

擬人化エージェント条件およびラップトップ条件それぞれの問題に対する、参加者 30 名全体での正解率を図 15 に示す。擬人化エージェント条件での正解率は、ラップトップ条件に比べて有意に高い ($p = 0.0003$) ことが認められる。また、A 条件および L 条件それぞれにおける、各問題に対する正解率を図 16 に示す。どちらの条件でも、擬人化エージェント条件での正解率がラップトップ条件に比べて有意に高い (A 条件: $p = 0.0320$, L 条件: $p = 0.0004$) ことが認められ、各問題間での正解率の差は L 条件の方が大きかった。加えて、参加者の正解率について、2 要因混合計画で分散分析を行なった。その結果、擬人化エージェント/ラップトップ要因の主効果が認められ ($F(1, 28) = 16.94, p = 0.0003, \text{partial } \eta^2 = 0.3769$)、順序要因の主効果は認められず ($F(1, 28) = 1.26, p = 0.2712, \text{partial } \eta^2 = 0.0431$)、順序要因と擬人化エージェント/ラップトップ要因間に交互作用は認められなかった ($F(1, 28) = 1.26, p = 0.2712, \text{partial } \eta^2 = 0.0431$)。これらのことから、参加者の正解率は問題の内容にのみ依存しており、問題を解く順序は影響しないと考えられる。

次に、参加者 30 名全体での、各問題で正解するまでに要した時間、すなわち経過時間の平均値を図 17 に示す。擬人化エージェント条件での経過時間は、ラップトップ条件に比べて有意に短い ($p = 0.0018$) こと

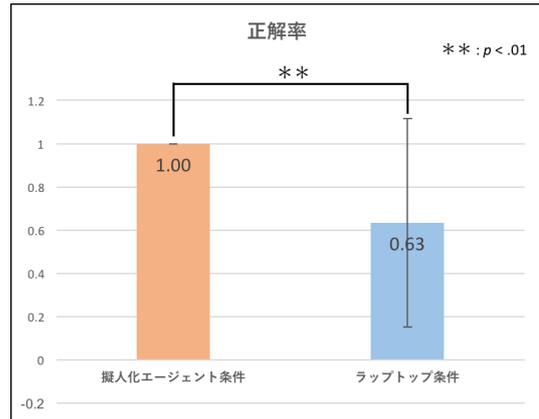


図 15: 参加者 30 名全体での各問題の正解率

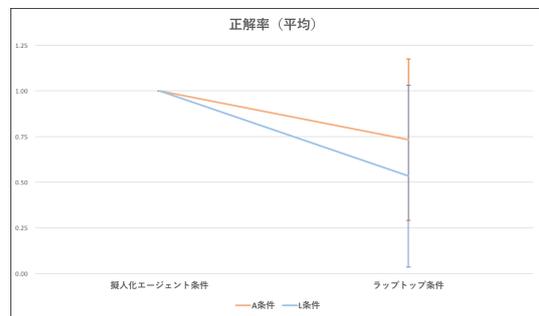


図 16: A 条件および L 条件での各問題の正解率

が認められる。また、A 条件および L 条件それぞれにおける、各問題での経過時間の平均値を図 18 に示す。A 条件では 2 つの問題間で経過時間に有意差は認められない ($p = 0.2660$) が、L 条件では擬人化エージェント条件の方が経過時間が有意に短い ($p = 0.0013$) ことが認められる。加えて、経過時間の差が最も大きくなる組合せは、A 条件における擬人化エージェント条件、および L 条件におけるラップトップ条件だった ($p = 0.0084$)。一方、経過時間について 2 要因混合計画で分散分析を行なった結果、擬人化エージェント/ラップトップ要因の主効果が認められ ($F(1, 28) = 12.80, p = 0.0013, \text{partial } \eta^2 = 0.3137$)、順序要因の主効果は認められず ($F(1, 28) = 1.63, p = 0.2122, \text{partial } \eta^2 = 0.0550$)、順序要因と擬人化エージェント/ラップトップ要因間に交互作用は認められなかった ($F(1, 28) = 3.53, p = 0.0707, \text{partial } \eta^2 = 0.1119$)。これらのことから、経過時間は問題の内容にのみ依存しており、問題を解く順序は影響しないと考えられる。

加えて、参加者 30 名全体での、各問題についての主観難易度の平均値を図 19 に示す。擬人化エージェント条件での難易度は、ラップトップ条件に比べて有意に低い ($p = 0.0001$) ことが認められる。また、A 条

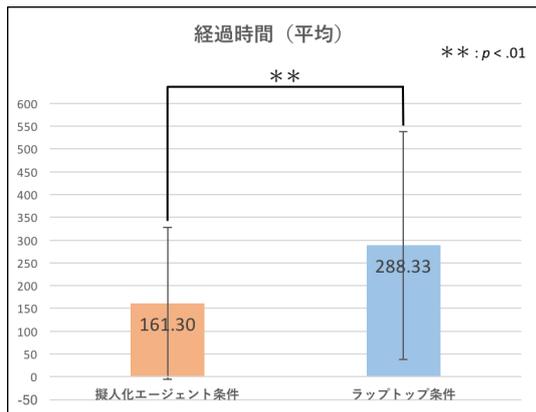


図 17: 参加者 30 名全体での各問題における経過時間の平均値

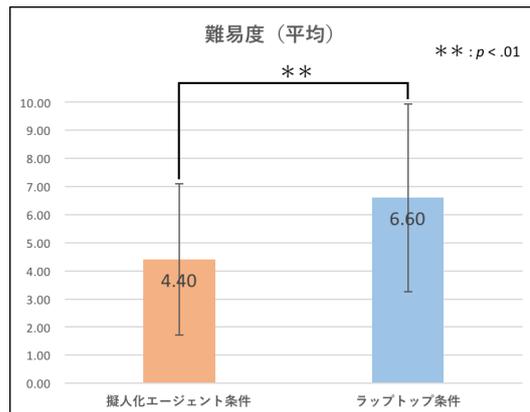


図 19: 参加者 30 名全体での各問題の主観難易度の平均値

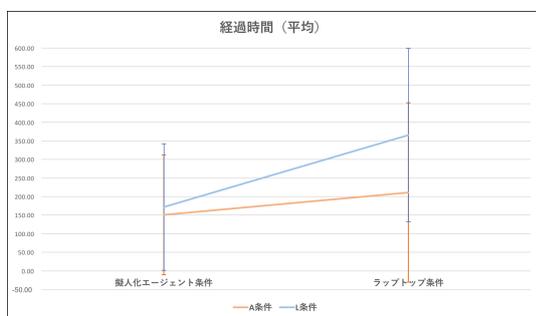


図 18: A 条件および L 条件での各問題における経過時間の平均値

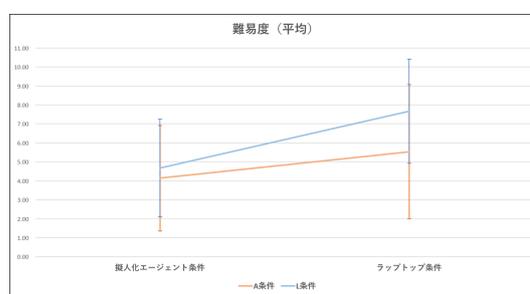


図 20: A 条件および L 条件での各問題の主観難易度の平均値

件および L 条件それぞれにおける、各問題での参加者の主観難易度の平均値を図 20 に示す。A 条件では擬人化エージェント条件の難易度はラップトップ条件に比べて有意に低い傾向にある ($p = 0.0638$) 程度だが、L 条件では擬人化エージェント条件の難易度が有意に低い ($p = 0.0004$) ことが認められる。加えて、難易度の差が最も大きくなる組合せは、A 条件における擬人化エージェント条件、および L 条件におけるラップトップ条件だった ($p = 0.0021$)。一方、主観難易度について 2 要因混合計画で分散分析を行なった結果、擬人化エージェント/ラップトップ要因の主効果が認められ ($F(1, 28) = 21.22, p = 0.0001, \text{partial } \eta^2 = 0.4311$)、順序要因の主効果は認められず ($F(1, 28) = 1.78, p = 0.1929, \text{partial } \eta^2 = 0.0597$)、順序要因と擬人化エージェント/ラップトップ要因間に交互作用は認められなかった ($F(1, 28) = 2.81, p = 0.1048, \text{partial } \eta^2 = 0.0911$)。これらの事から、主観難易度は問題の内容にのみ依存しており、問題を解く順序は影響しないと考えられる。

4.5.2 アンケート

表 2 に示す 15 項目の質問に対する、参加者 30 名全体での平均評価値を図 21 に示す。また、これらをもとに算出された、参加者 30 名全体での精神負担、身体負担、時間圧力、作業成績、フラストレーションの平均値を図 22 に示す。精神負担が特に大きく、参加者にとって本実験は「難しかった」「複雑だった」「頭を使う必要があった」「正確さが要求された」ことを示す。逆に、身体負担、時間圧力、作業成績、フラストレーションの 4 つは負の値を示しており、参加者はそれら 4 つに関する不満を感じていなかったことを示す。次に、A 条件および L 条件それぞれにおける、15 項目の質問に対する平均評価値を図 23 に示す。また、これらをもとに算出された、A 条件および L 条件における精神負担、身体負担、時間圧力、作業成績、フラストレーションの平均値を図 24 に示す。「楽しかった」「集中していた」の 2 項目が、L 条件の方が有意に大きい傾向にある ($p = 0.0661, p = 0.0713$) 程度で、その他では両条件間で有意差は認められなかった。

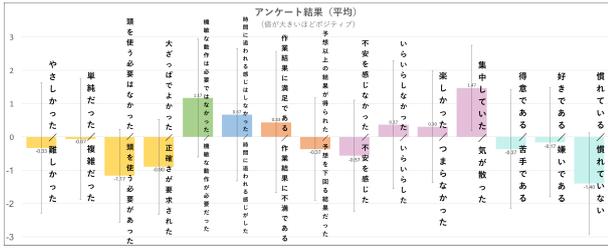


図 21: 15 項目の質問に対する参加者 30 名全体での平均評価値

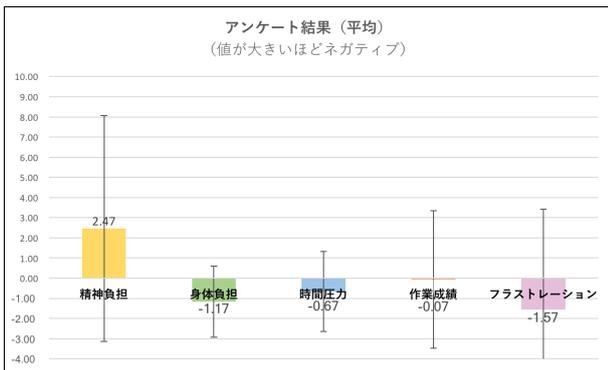


図 22: 参加者 30 名全体での精神負担, 身体負担, 時間圧力, 作業成績, フラストレーションの平均値

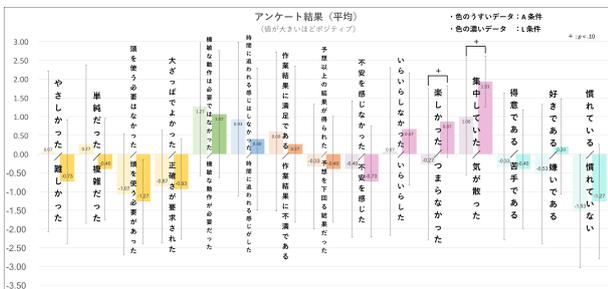


図 23: A 条件および L 条件における 15 項目の質問に対する平均評価値

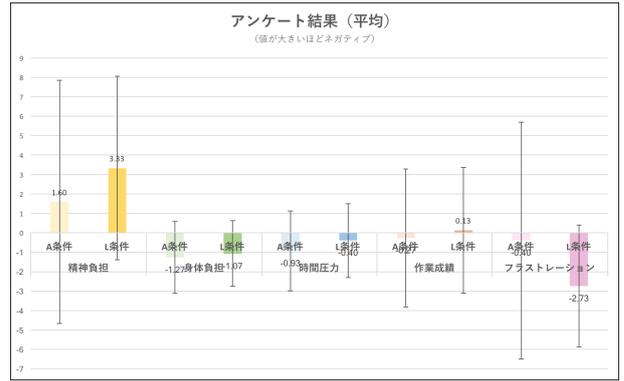


図 24: A 条件および L 条件における精神負担, 身体負担, 時間圧力, 作業成績, フラストレーションの平均値

次に、アンケートの Q1~Q4 に対する、参加者 30 名全員分の回答を図 25 に示す。Q1 より、擬人化エージェント条件の方がラップトップ条件よりやる気が出たと回答した参加者が 15 名で最も多い。一方、「いいえ」と回答した参加者は 3 名で最も少なく、「はい」あるいは「どちらでもない」と回答した参加者よりも有意に少ない (いずれも $p < 0.01$) と認められる。また、Q2 より、ラップトップ条件で画面に擬人化エージェントが表示された方がやる気が出ると回答した人数は 25 名で最多であり、「いいえ」あるいは「どちらでもない」と回答した人数より有意に多い (いずれも $p < 0.01$) と認められる。Q3 より、擬人化エージェント条件の方が日常的だと回答した人数は 23 名で最多であり、ラップトップ条件の方が日常的、あるいはどちらでもないと回答した人数より有意に多い (いずれも $p < 0.01$) と認められる。加えて、Q4 より、本実験で出題された状態遷移に規則性があると感じた人数は 24 名で、規則性があると感じなかった人数よりも有意に多い ($p < 0.01$) と認められる。次に、A 条件および L 条件それぞれにおける、Q1~Q4 への回答を図 26 に示す。Q1 より、擬人化エージェント条件の方がラップトップ条件よりやる気が出た人数は、L 条件の方が A 条件よりも有意に多い傾向にある ($p = 0.0719$) と認められる。また、A 条件では「どちらでもない」と回答した参加者が 8 名で最も多く、「はい」と回答した参加者よりも有意に多い傾向にあり ($p = 0.0824$)、「いいえ」と回答した参加者よりも有意に多い ($p = 0.0086$) と認められる。一方、L 条件では「はい」と回答した参加者が 10 名で最も多く、他の 2 種類の回答よりも回答数が有意に多い (いずれも $p < 0.01$) と認められる。Q2 より、A 条件および L 条件のどちらにおいても、ラップトップ条件で画面に擬人化エージェントが表示された方がやる気が出ると回答した人数が最多であり、他の 2 種類の回答よりも有意に多い (いずれも $p < 0.01$) と認められる。ただし、いずれの回答とも、両条件間で回答数に有意差は

認められなかった (いずれも $p > 0.10$)。また、Q3 より、A 条件および L 条件のどちらにおいても、擬人化エージェント条件の方が日常的だと回答した人数が最多であり、他の 2 種類の回答よりも有意に多い (いずれも $p < 0.05$ あるいは $p < 0.01$) と認められる。加えて、擬人化エージェント条件の方が日常的だと回答した人数は、L 条件の方が有意に多く ($p = 0.0341$)、ラップトップ条件の方が日常的だと回答した人数は、L 条件の方が有意に少ない ($p = 0.0320$) と認められる。Q4 より、A 条件および L 条件のどちらにおいても、本実験で出題された状態遷移に規則性があると感じた人数は、規則性を感じなかった人数よりも有意に多い (両条件とも $p < 0.01$) と認められる。また、各回答とも両条件間で回答数に有意差は認められなかった (いずれも $p > 0.10$)。さらに、Q4 について、どのような規則性があると感じたかを参加者に記述してもらった。参加者 30 名全員分の回答を図 27 に示す。回答は主に 6 種類に大別された。なかでもミーリ・マシンの一般的な性質に関する回答が 11 名で最も多く、他の 5 種類の回答よりも有意に多い (いずれも $p < 0.05$ あるいは $p < 0.01$) と認められる。加えて、A 条件および L 条件それぞれにおける、どのような規則性があると感じたかの回答を図 28 に示す。A 条件では、各回答とも回答数が 0~4 で大差はなく、いずれの組合せとも回答数に有意差は認められなかった (いずれも $p > 0.05$)。一方、L 条件ではミーリ・マシンの一般的な性質に関する回答が 7 名で、規則性があると感じなかった 4 名より有意に多い傾向にあり ($p = 0.0824$)、他の 4 種類の回答より有意に多い (いずれも $p = 0.0086$) と認められる。また、6 種類の回答全てにおいて、両条件間で回答数に有意差は認められなかった (いずれも $p > 0.10$)。

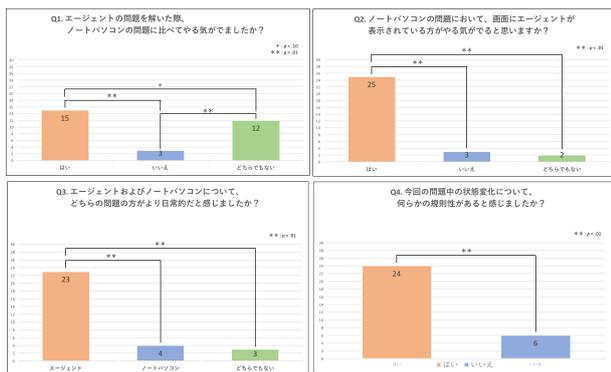


図 25: Q1~Q4 に対する参加者 30 名全員分の回答

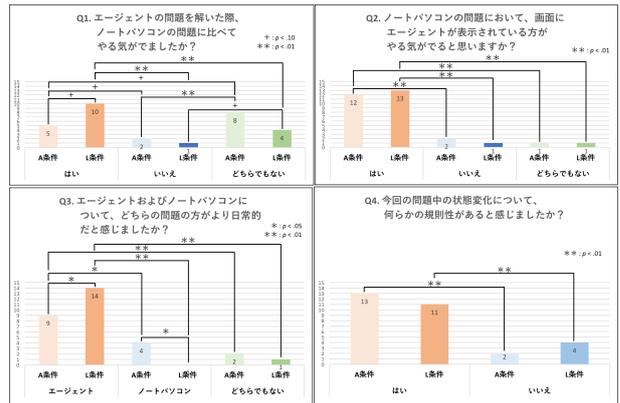


図 26: A 条件および L 条件における Q1~Q4 への回答

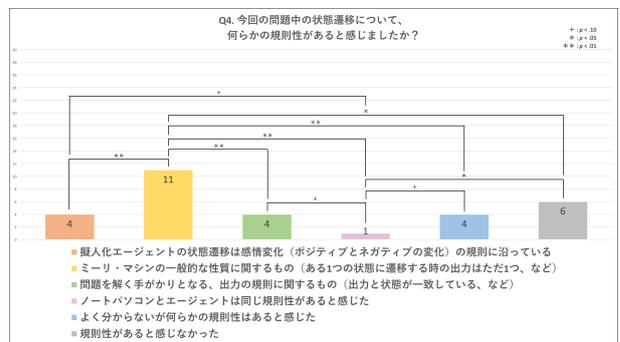


図 27: Q4 に対する参加者 30 名全員分の回答

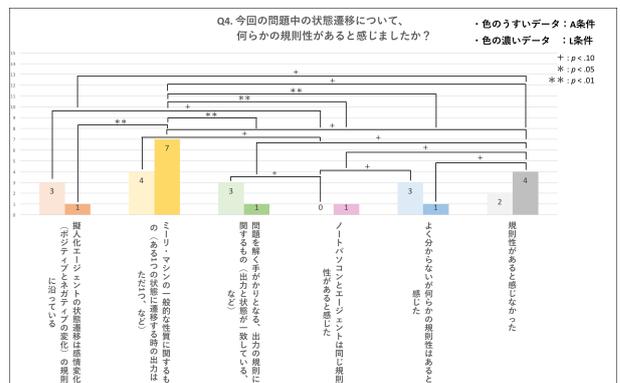


図 28: A 条件および L 条件における Q4 への回答

アンケートの Q5 に対する、参加者 30 名全員分の回答を図 29 に示す。回答は主に 4 種類に大別された。そのうち、擬人化エージェントの内部状態について、「感情、精神状態、気持ち」と回答した参加者が 19 名で最も多く、他の 3 種類の回答よりも回答数が有意に多い (いずれも $p < 0.01$) と認められる。他の 3 種類の回答については、回答数に有意差は認められなかった (いずれも $p > 0.05$)。また、A 条件および L 条件それぞれ

れにおける、Q5 への回答を図 30 に示す。両条件とも、「感情、精神状態、気持ち」と回答した参加者が最も多く、A 条件では他の 3 種類の回答よりも有意に多い（いずれも $p < 0.01$ ）と認められる。L 条件でも、「その他」に分類される回答をした人数よりも有意に多い傾向にあり（ $p = 0.0824$ ）、他の 2 種類の回答よりも有意に多い（いずれも $p < 0.01$ ）と認められる。また、「その他」に分類される回答をした参加者は、L 条件の方が A 条件よりも有意に多い（ $p = 0.0132$ ）と認められる。一方、他の 3 種類の回答については、両条件間で回答数に有意差は認められなかった（いずれも $p > 0.10$ ）。

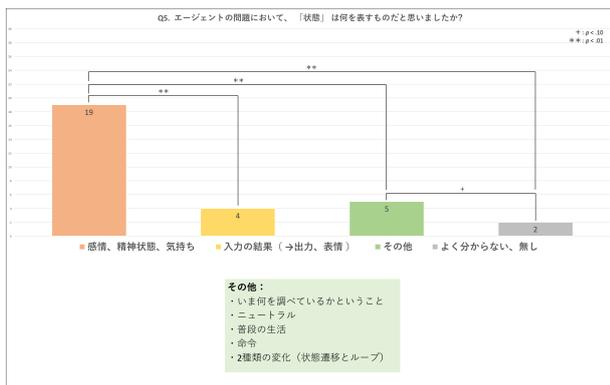


図 29: Q5 に対する参加者 30 名全員分の回答

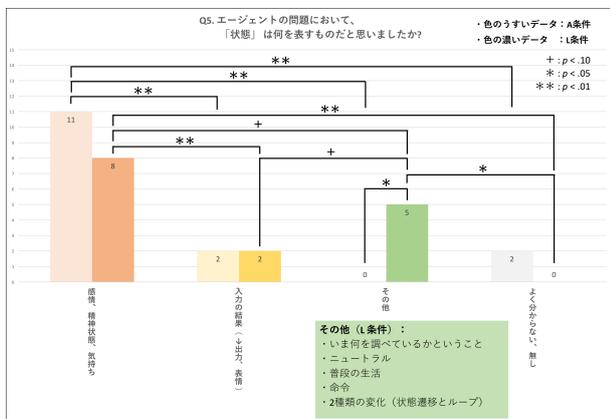


図 30: A 条件および L 条件における Q5 への回答

また、アンケートの Q6 に対する、参加者 30 名全員分の回答を図 31 に示す。回答は主に 6 種類に大別された。そのうち、ラップトップの内部状態について、「よく分からない、無し」と回答した参加者が 11 名で最も多く、他の 5 種類の回答よりも有意に多い（いずれも $p < 0.05$ あるいは $p < 0.01$ ）と認められる。また、A 条件および L 条件それぞれにおける、Q6 への回答を図 32 に示す。両条件とも、「よく分からない、無し」と回答した参加者が最多だった。ただし、L 条件では 6

種類の回答とも回答数は 1~4 で大差はなく、いずれの組合せとも回答数に有意差は認められなかった（いずれも $p > 0.05$ ）。一方、A 条件では「よく分からない、無し」と回答した参加者は、他の 5 種類の回答よりも有意に多い（いずれも $p < 0.05$ あるいは $p < 0.01$ ）と認められる。また、「入出力の結果」と回答した参加者は、L 条件の方が A 条件よりも有意に多い傾向にある（ $p = 0.0719$ ）と認められる。一方、その他の 5 種類の回答をした参加者の人数については、両条件間で有意差は認められなかった（いずれも $p > 0.10$ ）。

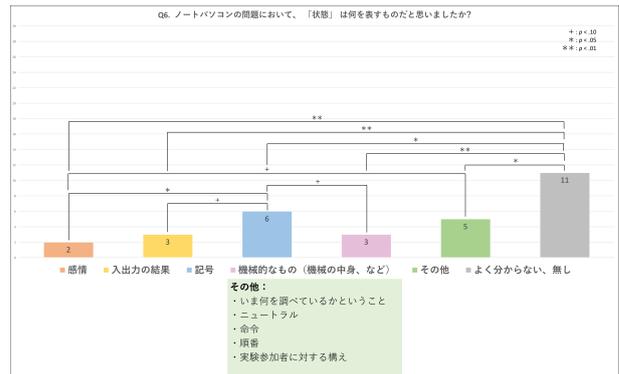


図 31: Q6 に対する参加者 30 名全員分の回答

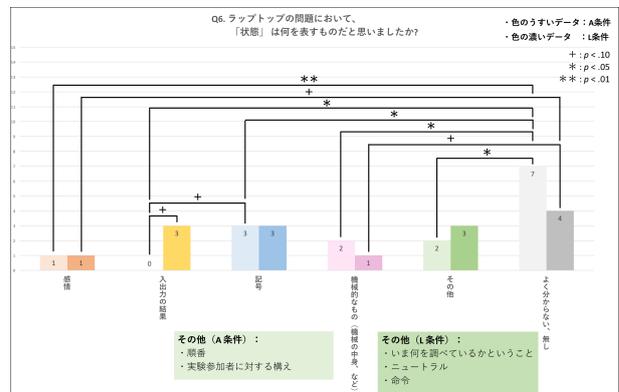


図 32: A 条件および L 条件における Q6 への回答

次に、アンケートの Q7 のうち、「エージェントの問題について、どのような点がやりやすかったですか？」という質問に対する、参加者 30 名全員分の回答を図 33 に示す。回答は主に 6 種類に大別された。そのうち、「入力と出力の関連性をイメージしやすい、規則性を理解しやすい」と回答した参加者が 18 名で最も多く、他の 5 種類の回答よりも有意に多い（いずれも $p < 0.01$ ）と認められる。他の 5 種類の回答は回答数が 2~4 で大差はなく、いずれの組合せとも回答数に有意差は認められなかった（いずれも $p > 0.10$ ）。また、A 条件および L 条件それぞれにおける、Q7 の同質問への回答を図 34

に示す。両条件とも、「入力と出力の関連性をイメージしやすい、規則性を理解しやすい」と回答した参加者が最多であり、他の5種類の回答よりも有意に多い(いずれも $p < 0.01$)と認められる。一方、他の5種類の回答については、両条件とも回答数が0~2で大差はなく、いずれの組合せとも回答数に有意差は認められなかった(いずれも $p > 0.10$)。また、6種類全ての回答について、両条件間で回答数に有意差は認められなかった(いずれも $p > 0.10$)。加えて、Q7のうち、「エージェントの問題について、どのような点が難しかったですか?」という質問に対する、参加者30名全員分の回答を図35に示す。回答は主に5種類に大別された。そのうち、「無し」と回答した参加者が14名で最も多く、他の4種類の回答よりも有意に多い(いずれも $p < 0.01$)と認められる。次いで「規則性を理解するのが難しかった」と回答した参加者が7名で2番目に多く、「慣れておらず戸惑った」と回答した人数よりも有意に多い傾向にあり($p = 0.0824$)、「表情の区別がつきにくかった」あるいは「その他」に分類される回答をした人数よりも有意に多い(いずれも $p < 0.05$)と認められる。また、A条件およびL条件それぞれにおける、Q7の同質問への回答を図36に示す。両条件とも、「無し」と回答した人数が7名で最多であり、A条件では「慣れておらず戸惑った」と回答した4名よりも有意に多い傾向にあり($p = 0.0824$)、他の3種類の回答よりも有意に多い(いずれも $p < 0.05$ あるいは $p < 0.01$)と認められる。また、L条件では「規則性を理解するのが難しかった」を除く他の3種類の回答よりも有意に多い(いずれも $p < 0.05$ あるいは $p < 0.01$)と認められる。また、「慣れておらず戸惑った」と回答した参加者は、A条件の方がL条件よりも有意に多い($p = 0.0320$)ことが認められる。一方、他の4種類の回答については、両条件間で回答数に有意差は認められなかった(いずれも $p > 0.10$)。

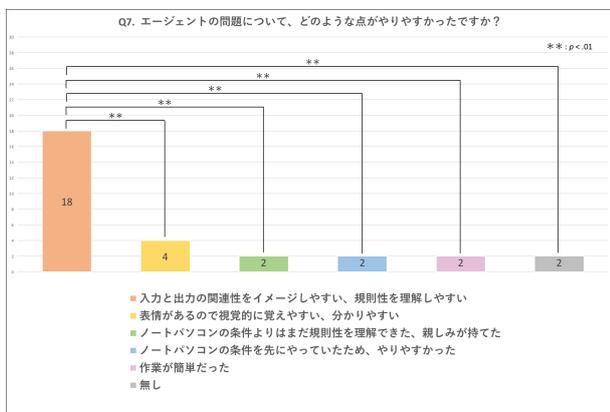


図 33: Q7 (やりやすかった点) に対する参加者 30 名全員分の回答

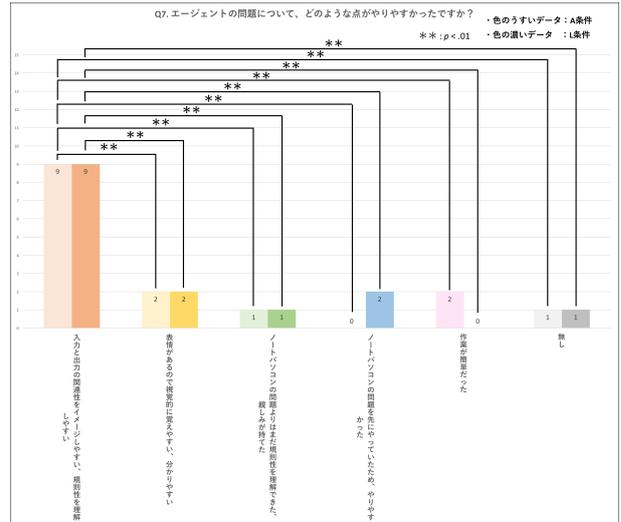


図 34: A 条件および L 条件における Q7 (やりやすかった点) への回答

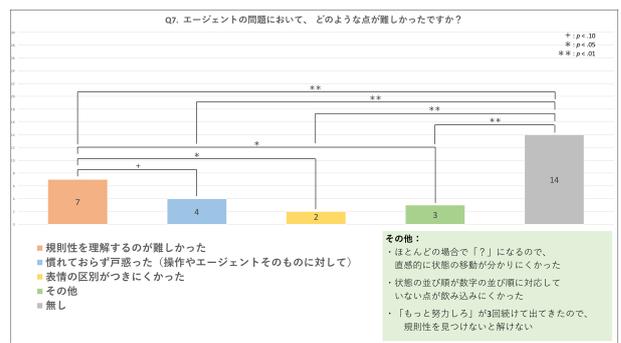


図 35: Q7 (難しかった点) に対する参加者 30 名全員分の回答

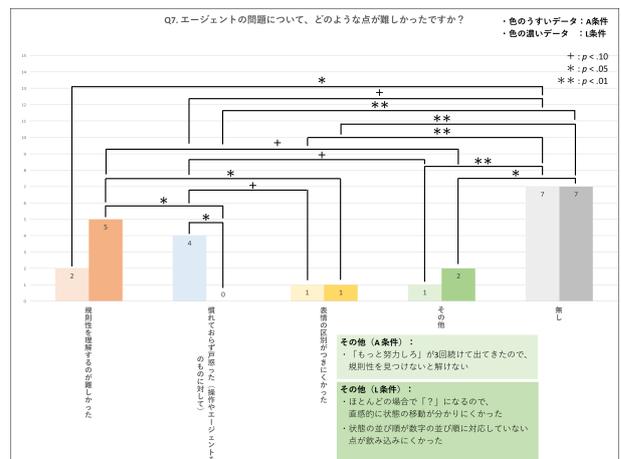


図 36: A 条件および L 条件における Q7 (難しかった点) への回答

また、アンケートのQ8のうち、「ノートパソコンの問題について、どのような点がやりやすかったですか？」という質問に対する、参加者30名全員分の回答を図37に示す。回答は主に6種類に大別された。そのうち、「無し」と回答した参加者が15名で最も多く、他の5種類の回答よりも有意に多い（いずれも $p < 0.01$ ）と認められる。一方、他の5種類の回答については、回答数が2~5で大差はなく、いずれの組合せとも回答数に有意差は認められなかった（いずれも $p > 0.05$ ）。また、A条件およびL条件それぞれにおける、Q8の同質問への回答を図38に示す。両条件とも「無し」と回答した人数が最多であり、A条件では「記号で考えられるので解きやすかった」あるいは「作業が単純だった」と回答した3名よりも有意に多い傾向にあり（いずれも $p = 0.0824$ ）、その他3種類の回答をした1名よりも有意に多い（いずれも $p = 0.0192$ ）と認められ、L条件では他の5種類の回答よりも有意に多い（いずれも $p < 0.01$ ）と認められる。また、「作業が単純だった」と回答した参加者は、A条件の方がL条件よりも有意に多い傾向にある（ $p = 0.0719$ ）と認められる。一方、他の5種類の回答については、両条件間で回答数に有意差は認められなかった（いずれも $p > 0.10$ ）。加えて、Q8のうち、「ノートパソコンの問題について、どのような点が難しかったですか？」という質問に対する、参加者30名全員分の回答を図39に示す。回答は主に5種類に大別された。そのうち、「規則性を理解するのが難しかった」と回答した参加者が20名で最も多く、他の4種類の回答よりも有意に多い（いずれも $p < 0.01$ ）と認められる。一方、他の4種類の回答については、回答数が1~4で大差はなく、いずれの組合せとも回答数に有意差は認められなかった（いずれも $p > 0.05$ ）。また、A条件およびL条件それぞれにおける、Q8の同質問への回答を図40に示す。両条件とも「規則性を理解するのが難しかった」と回答した人数が最多であり、他の4種類の回答よりも有意に多い（いずれも $p < 0.05$ あるいは $p < 0.01$ ）と認められる。また、「無し」と回答した参加者は、A条件の方がL条件よりも有意に多い傾向にある（ $p = 0.0719$ ）と認められる。一方、他の4種類の回答については、両条件間で回答数に有意差は認められなかった（いずれも $p > 0.10$ ）。

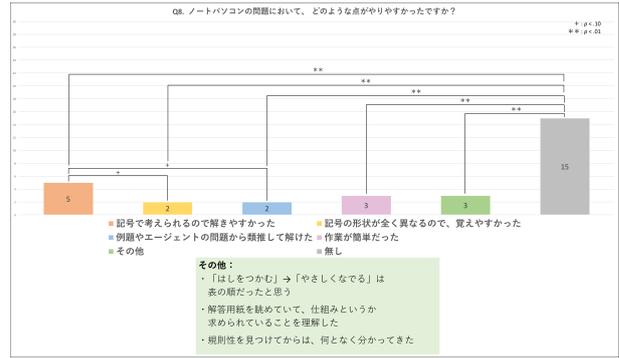


図 37: Q8（やりやすかった点）に対する参加者30名全員分の回答

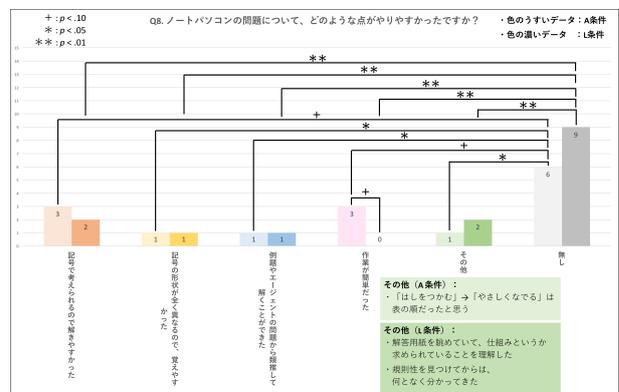


図 38: A条件およびL条件におけるQ8（やりやすかった点）への回答

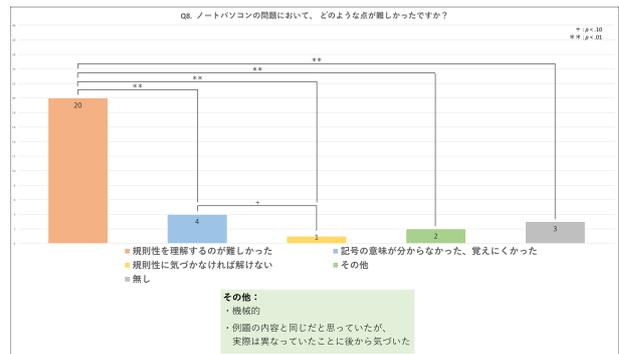


図 39: Q8（難しかった点）に対する参加者30名全員分の回答

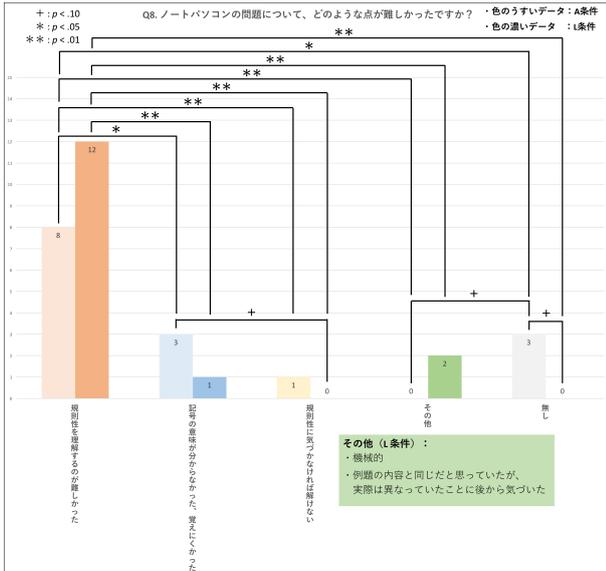


図 40: A 条件および L 条件における Q8（難しかった点）への回答

5 考察

本章では、4.5 節に示した結果を考察する。

5.1 各エージェント毎のパフォーマンスの比較

本実験における参加者 30 人全体では、正解率、経過時間、主観難易度の全てにおいて、擬人化エージェント条件とラップトップ条件とで有意差が認められ、擬人化エージェント条件の方が参加者のパフォーマンスが優れていた。擬人化エージェント条件では、インタラクションの内容が実世界との類似度が高いものとして設定されている。そのため、参加者が擬人化エージェントのミーリー・マシンを学習する際に認知バイアスが生じ、結果的にラップトップ条件に比べて学習速度が速く、推定パフォーマンスが優れていたと考えられる。すなわち、本実験の結果は、インタラクションの内容の違いに応じた認知バイアスが生じ、人間の推定パフォーマンスに変化が生じるという仮説を支持するものと考えられる。

また、A 条件と L 条件とで、擬人化エージェント条件とラップトップ条件間での参加者のパフォーマンスの差に違いが見られた。まず、擬人化エージェント条件については、正解率、経過時間、主観難易度の全てにおいて、A 条件と L 条件とで差は見られなかった。また、A 条件では、擬人化エージェント条件の方がラップトップ条件より正解率が有意に高いことが認められるが、経過時間および主観難易度については両条件間で有意差は認められなかった。一方、L 条件では、正

解率、経過時間、主観難易度の全てにおいて、擬人化エージェント条件の方がラップトップ条件より参加者の成績が有意に優れていた。これに対し、2 要因混合計画で分散分析を行なった結果より、正解率、経過時間、主観難易度はいずれも問題の内容にのみ依存しており、問題を解く順序は影響しないと考えられる。以上より、インタラクションの内容の違いに応じた認知バイアスが生じ、人間の推定パフォーマンスに変化が生じるという仮説を支持する結果が得られたと考えられる。

5.2 アンケートの評価

表 2 に示す 15 項目の質問に対する、参加者 30 名全体での平均評価値より、本実験におけるタスクは参加者にとって不慣れかつ困難なものであったと考えられる。表 24 に示す精神負担の値も、そのことを示していると考えられる。また、表 2 の 15 項目の質問に対する、A 条件および L 条件での平均評価値は、問題を解く順番によって参加者の主観的ワークロードの評価値に差が生じないことを示す。このことから、本実験における参加者のパフォーマンスの変化は、参加者の主観的ワークロードの評価値の変化によるものではないと考えられる。

次に、アンケートの Q1 への回答結果より、参加者は L 条件でのみ、擬人化エージェント条件でラップトップ条件よりもやる気が出たと言える。しかし、擬人化エージェント条件における参加者のパフォーマンスについて、A 条件と L 条件とでは有意差は認められなかった。仮に、参加者の「やる気」の有無がパフォーマンスに影響するのであれば、擬人化エージェント条件での参加者のパフォーマンスは、L 条件の方が有意に優れると考えられる。これらのことから、参加者の「やる気」の有無はパフォーマンスに影響しないと考えられる。また、Q3 への回答結果より、問題が日常的だと感じるかどうかは問題の内容にのみ依存しており、問題を解く順序は影響しないと考えられる。これに対し、Q1 より擬人化エージェント条件に対して参加者のやる気が出たのは L 条件だけであり、これらのことから、参加者のパフォーマンスに影響するのは、参加者の「やる気」ではなく「問題の日常性」であると考えられる。また、Q1 で回答された「やる気」とは、参加者の「各問題の成績に対する主観評価値」を反映したものではないかと考えられる。A 条件では、経過時間および主観難易度について擬人化エージェント条件とラップトップ条件とで有意差は認められず、結果的に Q1 で「どちらでもない」と回答した人数が最も多かった。一方、L 条件では、正解率、経過時間、主観難易度の全てにおいて、擬人化エージェント条件の方がラップトップ条件より有意に優れていると認められ、結果的に Q1 で「は

い」すなわち擬人化エージェント条件の方がやる気が出たと回答した人数が最も多かった。これらのことから、Q1で回答された「やる気」とは、参加者の「各問題の成績に対する主観評価値」を反映したものである可能性が示唆される。一方、Q2について、各回答の人数の分布がA条件とL条件とでほぼ一致しており、A条件とL条件のどちらも擬人化エージェントがラップトップの画面上に表示された方がやる気が出ると回答した人数が最多である点から、Q2で回答された「やる気」は、Q3で回答された「問題の日常性」と同様のものを示していると考えられる。

また、Q4について、A条件では出題されたミーリ・マシンに対する、参加者全体で共通して見出された「規則性」は存在しなかったと考えられる。一方、L条件では「ミーリ・マシンの一般的な性質に関するもの」についての回答をした人数が最も多く、他の5種類の回答よりも有意に多い傾向にある、あるいは有意に多いと認められる。しかし、「ミーリ・マシンの一般的な性質に関するもの」とは、「ある1つの状態に遷移する際の出力はただ1つである」、「ある状態において入力を与えられると、出力をして別の状態に遷移する、あるいは同じ状態をループする」などのミーリ・マシンの定義に関する回答であり、本実験で出題された問題を解く上での直接的な手掛かりになり得るものではない。また、6種類の回答いずれとも、A条件とL条件とで回答数に有意差は認められなかった。これらのことから、本実験においては、問題を解く上での直接的な手掛かりになり得るような、参加者全体で共通して見出された「規則性」は存在しなかったと考えられる。あるいは、そのような「規則性」は存在したものの、Q4には回答されなかった可能性も考えられる。

次に、Q5への回答結果より、参加者が擬人化エージェント条件の問題を解いた際、その内部状態について、参加者全体で共通して見出された「意味」（今回の場合は「感情、精神状態、気持ち」）が存在したと考えられる。一方、Q6への回答結果より、参加者はラップトップ条件の問題を解いた際、その内部状態について、「意味」を見出せなかったと考えられる。これらのことから、内部状態の「意味」を発見するプロセスの有無が、参加者のパフォーマンスに影響したと考えられる。擬人化エージェント条件では、インタラクションの内容が実世界との類似度が高いものとして設定されており、そのおかげで内部状態の「意味」を見出すことができたと考えられる。この「意味」の発見のプロセスが、認知バイアスによるものである可能性が示唆される。

また、Q7への回答結果より、参加者は擬人化エージェント条件において、規則性を理解しやすく、一方で難しさをあまり感じなかった。これは、インタラクションの内容が実世界との類似度が高いものとして設定されていたためだと考えられる。一方、Q8への回答

結果より、参加者はラップトップ条件において、規則性を理解するのが難しく、やりやすさをあまり感じなかった。これは、インタラクションの内容が実世界との類似度が低いものとして設定されていたためだと考えられる。これらのことから、ミーリ・マシン学習の難易度は、インタラクションの内容の実世界との類似度にも影響されると考えられる。この結果は、人間がミーリ・マシンを学習する際には、インタラクションの内容の違いに応じた認知バイアスが生じ、人間の推定パフォーマンスに変化が生じるという仮説を支持するものと考えられる。

6 むすび

本研究では、人間がエージェントとインタラクションを行い、そのエージェントの内部状態遷移をミーリ・マシンとして推定する状況を対象とし、インタラクションの内容が異なる場合に人間の推定パフォーマンスがどのように変化するかを実験的に解明することを目的とした。本実験では、独立変数として擬人的特性の異なる2種類のエージェントを設定し、それらのエージェントに基づき、実世界との類似度が異なる2種類のインタラクション場面を構築し、それらの差異が人間の推定パフォーマンスに与える影響、すなわち人間の認知バイアスとの関係性に注目した。本実験では、擬人化エージェントおよびラップトップの2種類のエージェントを使用し、擬人化エージェントとは実世界との類似度が高いインタラクション場面を、ラップトップとは実世界との類似度が低いインタラクション場面をそれぞれ構築した。本実験では、擬人化エージェント条件およびラップトップ条件の問題に対する、参加者の正解率、参加者が正解のミーリ・マシンを学習するまでの経過時間（すなわち学習速度）、および各問題についての参加者の主観難易度のデータを収集し、これらを参加者のパフォーマンスとして計測した。また、各エージェント条件の問題が全て終わった後、参加者にアンケートへの回答をしてもらった。本実験では、先に擬人化エージェント条件の問題を解く場合（A条件）と、先にラップトップ条件の問題を解く場合（L条件）とで、参加者を同数ずつ2通りに分けた。

実験結果より、参加者全体での平均値では、正解率、経過時間、主観難易度の全てにおいて、擬人化エージェント条件とラップトップ条件とで有意差が認められ、擬人化エージェント条件の方が参加者のパフォーマンスが優れていた。また、2要因混合計画で分散分析を行なった結果、正解率、経過時間、主観難易度はいずれも問題の内容にのみ依存しており、問題を解く順序は影響しないと考えられる。加えて、アンケート結果より、擬人化エージェント条件でのみ、参加者はミー

リ・マシンの内部状態の「意味」を見出しており、規則性を理解しやすく、問題を解く上で難しさを感じていなかったと考えられる。これらのことから、ミーリ・マシン学習の難易度は、インタラクションの内容の実世界との類似度にも影響されると考えられる。擬人化エージェント条件では、インタラクションの内容が実世界との類似度が高いものとして設定されており、そのおかげで内部状態の「意味」を見出すことができたと考えられる。この「意味」の発見のプロセスが、認知バイアスによるものである可能性が示唆される。すなわち、参加者が擬人化エージェントのミーリ・マシンを学習する際に認知バイアスが生じ、結果的にラップトップ条件に比べて学習速度が速く、推定パフォーマンスが優れていたと考えられる。

以上より、人間がエージェントとインタラクションを行い、そのエージェントの内部状態遷移をミーリ・マシンとして推定する際、エージェントの擬人的特性の差異を含む、インタラクションの内容の違いに応じた認知バイアスが生じ、人間の推定パフォーマンスに変化が生じるという仮説が実証された。

参考文献

- [1] Kahneman, D., Slovic, P., Tversky, A. Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press, 1982.
- [2] George H. Mealy. A method for synthesizing sequential circuits. The Bell System technical journal, vol. 34, pp.1045-1079, 1955.
- [3] Cosmides, L. and Tooby, J. Cognitive Adaptations for Social Exchange. In: Barkow, J.H., Cosmides, L., Tooby, J., Eds., The Adapted Mind: Evolutionary Psychology and the Generation of Culture, Oxford University Press, Oxford, pp.163-228, 1992.
- [4] Kazunori Terada, Seiji Yamada, Akira Ito: Experimental Investigation of Human Adaptation to Change in Agent's Strategy through a Competitive Two-Player Game, In Proceedings of the 30th Annual CHI Conference on Human Factors in Computing Systems (CHI 2012), pp.2807-2810, Austin, USA, May 2012.
- [5] Kazunori Terada and Seiji Yamada: Mind-Reading and Behavior-Reading against Agents with and without Anthropomorphic Features in a Competitive Situation, *frontiers in Psychology*, July 2017.
- [6] Angluin, D., Learning regular sets from queries and counterexamples, *Inf. Comput*, 75 (2), pp.87-106, 1987.
- [7] Isberner, M., Foundations of active automata learning: An algorithmic perspective, PhD thesis, Technical University of Dortmund, 2015.
- [8] Isberner, M., Howar, F., Steffen, B., The TTT algorithm: A redundancy-free approach to active automata learning, In *RV '14, LNCS 8734*, Springer, pp.307-322, 2014.
- [9] Lee, D., Yannakakis, M., Principles and methods of testing finite state machines—A survey, *Proc. IEEE* 84 (8), pp.1090-1123., 1996.
- [10] Plous, S., *The Psychology of Judgment and Decision Making*, New York: McGraw-Hill, pp.233, 1993.
- [11] Wason, P, "Reasoning", In B. M. Foss(Ed.), *New horizons in psychology*, Harmondsworth: Penguin, 1966.
- [12] Cosmides, L.; Tooby, J., "Cognitive Adaptations for Social Exchange", In Barkow, J., *The adapted mind: Evolutionary psychology and the generation of culture*, New York: Oxford University Press, pp.163-228, 1992.
- [13] T. D. Bui, D. Heylen, M. Poel, and A. Nijholt, Generation of Facial Expressions from Emotion Using a Fuzzy Rule Based System, In *Proceedings of the 14th Australian Joint Conference on Artificial intelligence (AI 2001)*, Adelaide, Australia, pp.83-94, 2001.