

ロボットの振舞いの複雑さが 意図性認知に与える影響

今村 悠人^{1,a)} 寺田 和憲^{1,b)} 高橋 英之^{2,c)}

概要：心を持つロボットを作ることは人工知能研究者、ロボット研究者の目標であるがいまだにそのようなロボットは実現されていない。我々は心を「非ゼロ和ゲーム的状况において協調を実現するための手段として進化的に獲得されたエージェントの入出力関係の抽象表現」と考える。そして、入出力関係の抽象粒度が大きいことが心の本質であると考え、本研究で「人は入出力関係の抽象度を高くしなければ理解が困難な振舞いから心を感じる」という仮説を検証した。仮説の検証は機能発見能力、抽象粒度の選好、予測違反に対する理解、主観評定を指標として、ロボットの入出力関係が1対1の条件と1対多の条件を比較することで行った。実験の結果、人はロボットの振る舞いが「単純な入出力関係」よりも「抽象的で複雑な入出力関係」の場合に意図性を感じる可能性が高いことが示された。また、抽象化粒度の選好が心的状態の帰属に関係している可能性があることが示された。

キーワード：心、意図、ロボット、ヒューマンロボットインタラクション

1. はじめに

心を持つロボットを作ることは人工知能研究者、ロボット研究者の目標である。現在、産業用ロボットから単機能な家庭用ロボット、人に酷似したヒューマノイドロボットまで様々なロボットが開発されているが、未だに心を持ったと言い切れるロボットは存在しない。また、チェスや将棋のゲームにおいて人の能力を上回るコンピュータプログラムが開発されつつあるが、そのようなプログラムが心を持っていると言えるだろうか。

心とは何か。「心とは温かさを感じる神聖で何か特別なものである」という説明は間違っていないが客観性に欠ける。心とは人が進化の過程において生物としての究極因である自己保存・自己複製を達成するために獲得した機能である [1]。特に森林を追われた人類が開けたサバンナにおいて外敵の攻撃を防御するための手段として集団を形成するようになったことによって発生した社会の複雑さが心の進化に寄与したと考えられている [2]。

我々は心を「非ゼロ和ゲーム的状况において協調を実現するための手段として進化的に獲得されたエージェントの

入出力関係の抽象表現」と考える。醜いアヒルの子の定理によると事物の抽象化の軸は無限に存在する。その中で人は特定の基準を用いて事物の抽象化を行ってきた。この基準の選択は物理環境を反映しているのではなく、社会環境を反映したものであると考えられる。心、そして心のインスタンスである意図は漠然としている。すなわち抽象度が高い。抽象度を高くしなければならない理由は社会の複雑さにあると考えられる。構成個体が多い社会においてエージェント間の関係が複雑になるのは、エージェントが自己利益の追求と協力という相反する目標を追求しなければならないときである [3]。そして、社会において長期的な協力関係を維持するためには長い目で見た互恵的關係を築く必要がある。すなわち目先の利益に囚われたケチなエージェントは嫌われる。我々は、この自己利益の逸失に目をつぶり利他行動を可能しなければならないという制約が心を抽象度の高い概念、すなわちとらえどころのない曖昧なものにしている原因であると考えられる。

エージェントの入出力関係の抽象化粒度を大きくすることによって次のことが可能になる。

- (1) 個体間、個体内の両方において、代替手段の許容による問題解決を可能にする
- (2) 合理性を仮定することで他者の行動の将来の予測を可能にする（理解した気になる）

当事者研究を行っている自閉症スペクトラム障害 (ASD)

¹ 岐阜大学

² 大阪大学

a) yuto@elf.info.gifu-u.ac.jp

b) terada@gifu-u.ac.jp

c) hideyuki@ams.eng.osaka-u.ac.jp

者の綾屋らは ASD 者のかかえる認知的困難さを「まとめあげ困難」と表現している [4]。綾屋は著書で「おなかがすいた」こと「食べたいもの」が分からないと書いている。そのかわり、「胃のあたりが凹む」「胸が締まる」などの内蔵などの身体内部の状態変化に対する個別の感覚として知覚され、「冷たいもの」「もたれない」ものなど身体内部から発せられる詳細な要求項目が感じされると記している。そして、それらに合致するものは限定されているので、「ビビン麺」のような要求に合致するものが存在すれば飛びつくがそれ以外の場合はメニュー選択に苦勞すると記している。それに対して定型発達者 (TD) の抽象化については「人びとの『おなかがすいた』へのまとめあげは、確かにスピードは速いが実はとても大雑把で、うっかりしていることの裏返しではないだろうか」と書いている。TD 者は「おなかがすいた」という粒度の大きい抽象化を行うことで「とりあえず」目前のうどん屋に入るという意思決定、すなわち代替手段の許容による問題解決が可能になる。

入出力関係の抽象表現に曖昧性を含ませると他者の行動予測を容易にすることができる [5], [6], [7]。例えば、他者の目的を「飛行機に乗って東京に行く」のように比較的小さい粒度で抽象化を行っていると同風が来て飛行機が欠航した場合「飛行機が飛ばないので東京に行かない」という予測になる。しかし「東京に行く」という抽象化を行っていれば「新幹線に乗って東京に行く」という予測が可能になる。

以上のことから「人は入出力関係の抽象度を高くしなければ理解が困難な振舞いから心を感じる」という仮説が導かれる。そこで、本研究ではロボットの振舞いの複雑さが意図性認知に与える影響について調べた。我々はロボットに「機械的で単純な振る舞い」と「意図を持つ抽象的で複雑な振る舞い」の入出力関係を持たせ、異なる振る舞いの抽象度から受ける印象について調査を行った。本研究で用いたロボットの振る舞いの入出力関係は「1 対 1 関係 (図 1(a))」と「1 対多関係 (図 1(b))」の 2 種類である。1 対 1 関係の場合にはロボットはボタンスイッチを押すと必ず同一の行動を出力する。しかし 1 対多関係の場合には、ボタンスイッチに対して複数の行動が割り当てられており、一見ランダムに出力されているかのように見える。行動出力系列に対して目的を帰属することができれば、ロボットの振舞いを理解することができる。仮説の検証は機能発見能力、抽象粒度の選好、予測違反に対する理解、主観評定を指標として 1 対 1 条件と 1 対多条件を比較することで行った。

2. 実験方法

2.1 実験参加者

19 歳から 23 歳 (SD=1.05) の男性 10 人、女性 12 人の計 22 人で参加者間実験を行った。後述するが、実験 1 では男性 5 人女性 5 人、実験 2 では男性 5 人女性 7 人が実験

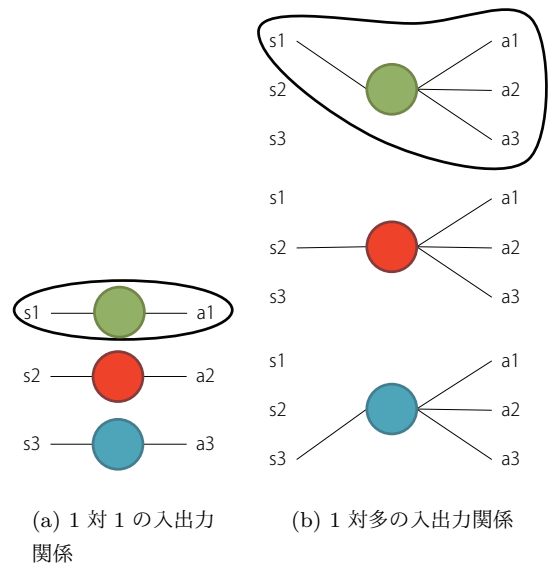


図 1 入出力関係



図 2 ロボットとリモートコントローラ

に参加した。

2.2 実験装置

ここでロボットの機能等の説明を記述する。図 2 に実際のロボットの写真を示す。ロボットはリバスト社の「マーキュリー・ロボット・キット 3WD13」を使用した。このロボットの直径は 310mm、高さは 230mm であり、周囲に 3 つの超音波センサが搭載されている。我々はこのロボットに次の 3 つの機能を追加して使用。

- 実験参加者とロボットの相互作用の際、実験参加者からの入力を受け付けるため赤、青、緑色のボタン各 1 つずつ設置した。
- ロボットの内部状態を表すために、ロボットの上部中央に ABS 樹脂製の半円型ドームで覆ったフルカラー LED を搭載した。フルカラー LED は光の三原色である赤、青、緑色の光を発することが出来る。
- ロボットを無線通信を用いて動かすために 3D プリンタで作成したリモートコントローラを用意した。このコントローラにはロボット上面と同じ赤、青、緑色のボタンが設置されている。無線通信には Digi 社の

XBee を用いている。

ロボットは次のように動作する。ユーザはロボットに設置されているボタンスイッチやリモートコントローラに設置されているボタンスイッチで押すことで入力を行う。その後モータが動作し、ロボットに付いているオムニホイールが動くことで行動単位を出力する。ボタンスイッチを押されたときに出力される一回の行動単位は、約 16.4cm である。またロボットの動作する行動の方向は赤、青、緑色のボタンスイッチの 3 種類の方向である。

このロボットは入出力の対応関係の複雑性 (抽象度) が可変である。入出力の対応関係は「1 対 1 関係」、「1 対多関係」の 2 段階である。ロボットに設置されているフルカラー LED は入力と出力の関係を表象している。発光は行動出力時のみ行う。この 2 段階ある対応関係とフルカラー LED によって表象する入出力関係の複雑さは以下の通りである。

- 1 対 1 関係

フルカラー LED が入力 1-出力 1 の関係を表象している。ロボットはボタンスイッチの種類に対応した動きを出力する。このときスイッチと動きは固定関係にある。フルカラー LED の発光色は入力のボタンスイッチの色を表象し、また行動の出力も表象する。

- 1 対多関係

フルカラー LED が入力 1-出力 3 の関係を表象している。ロボットはゴールに対して最適な行動 (ゴール到達行動) を出力する。例えばボタンスイッチ 1 を押すとゴール 1 に向かうような最適行動を出力する。このときフルカラー LED は入力したボタンスイッチの色を表象し、またスイッチに対応した「ゴール到達行動」を表象する。

その時々により出力される最適行動は、あらかじめ強化学習によって求めた行動価値関数に従って出力した。行動価値関数の学習は図 3 のように、30 (縦) × 30 (横) × 30 (角度) に離散化した環境 (x, y, θ) においてシミュレーションを行うことで行った。

3 輪オムニホイール型ロボットの運動学は付録に記載されている式 (A.1) を用いた。なお、ロボットのオムニホイールの半径が r 、各オムニホイールの角速度を $\omega_1, \omega_2, \omega_3$ 、中心からモータまでの距離を R とする。

状態 s' のロボットが行動 a を選び状態が s に遷移した時、 Q 値は次のようになる。

$$Q(s, a) = \sum_{s'} P(s'|a, s) (R(s, a, s') + \gamma \max_a Q(s', a)) \quad (1)$$

で求められる。この式を用いて、各ゴールに報酬を設定し、ロボットの位置と姿勢から各ゴールに対する最適行動を求めた。

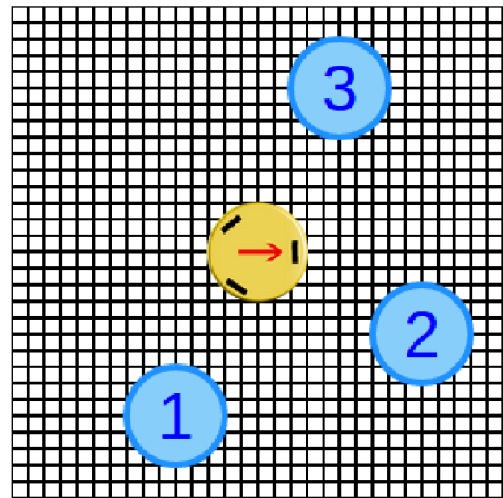


図 3 シミュレータ上の環境。この場合のロボットの状態は $(x, y, \theta) = (15, 15, 0)$

実験の環境を図 4 に示す。環境の広さは 3m × 3m 程度の正方形の環境である。実験環境には四角に囲まれた数字の 1, 2, 3 の合計 3 つのゴールを設置した。実験参加者にはリモートコントローラのボタンを押すことでロボットを操作し、ゴールへ到達させる「ゴール誘導タスク」を行ってもらった。

ロボットの位置や姿勢は環境上部に取り付けた USB カメラによって取得した。図 5 のように、USB カメラより実験環境全体を真上から撮影し (図 5 参照)、そこから得られた画像データに対して、OpenCV によるテンプレートマッチングを行うことでロボットの位置を検出し座標を取得した。

取得したロボットの座標と角度のデータを用いて、予め求めておいた行動価値関数からその場からゴールに向かうための最適行動を計算し、その結果をロボットへ送信した。ロボットはデータを受信し動作出力した。なお、PC-ロボット間の通信は Digi 社の「XStick」を用いて無線で行った。

ロボットの動作に「予測違反モード」を設けた。これはロボットが実験参加者の予測と異なる動作を行うモードである。例えば 1 対 1 モードの場合、当初スイッチ 1 に対してアクション 1 が対応付けられていたものが、突然スイッチ 1 に対してアクション 2 が対応付けられるようになるというものである。1 対多モードの場合、途中までスイッチ 1 を押したらゴール 1 に向かうような最適行動を行っていたが、途中からスイッチ 1 を押すとゴール 2 へ向かうようになるといったものである。予測違反モードでは本実験開始後、ボタンスイッチが 10 回押された時に対応付けの変化が起こるようにした。

2.3 測定方法

予測違反動作に対してどのような印象を持ったかをアン

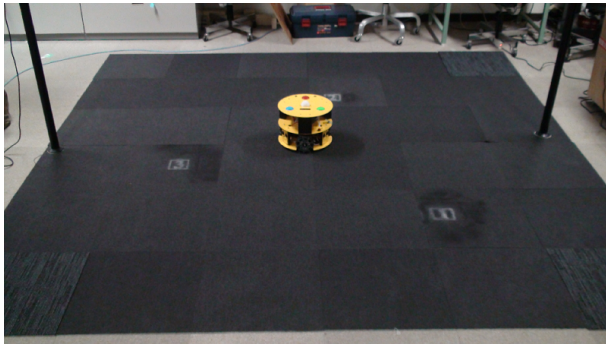


図 4 実験環境

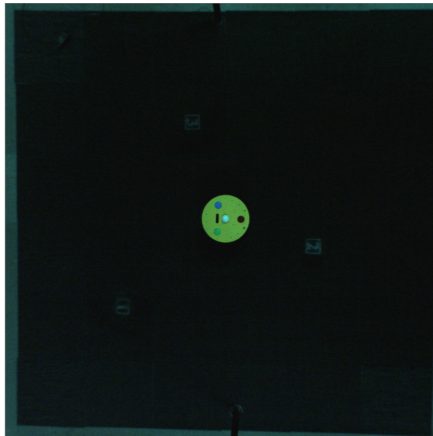


図 5 カメラから撮影した環境図

ケートによって測定した。アンケートは7段階のリッカート尺度を用いた。評価が7に近づくほど強い同意を示し、となり1に近づくほど強い不同意を示す。アンケートは表1の通りである。

表 1 実験後アンケート

質問番号	質問内容
1	驚いた
2	不思議に思った
3	疑問に感じた
4	壊れたと感じた
5	動きのルールが変わったと感じた
6	命令を拒否していると感じた
7	裏切られたと感じた
8	意表を疲れたと感じた
9	動きを予測できた
10	どんな感情を抱いたか (自由記述)

2.4 手順

2.4.1 実験 1

実験参加者にはロボットを使った簡単なゲームをする実験であると説明した。その後、ゲームの目的、ロボットの操作方法、モードが2種類存在する事など、一連の実験の流れを説明した。このとき、ロボットがどのように動くか

に関しては説明をせず、ボタンを押してロボットを動かしゴールへ向かわせるといった説明を行った。また、1対1、1対多モードを伝える際は「モード A」、「モード B」というようにして参加者にモードが2つあることを伝えた。そして一連の実験の流れとして、ロボットの動作を理解するために最初に練習として各モードを操作してもらうこと、本番前にどちらかモードを選択してもらうこと、そのモードで本番としてゴール誘導タスクを行ってもらうことを説明した。

まず最初にロボットの動作を理解してもらうため、参加者に対し一つのモードにつき10分間ロボットを操作してもらった。この時のロボットの操作方法はリモートコントローラを用いた間接的な操作であった。時間内に実験参加者がロボットの動きを理解したらロボットの動きを説明してもらい、参加者がロボットの動作を理解していたらその時点でそのモードの練習は終了した。その後、確認のため再度実験者がモードの説明を行った。参加者が説明した動作のルールが間違っていたならば、もう一度操作してもらい動作の理解をしてもらった。時間内に理解ができなかった場合、練習を止め、実験者がロボットの動きについて説明を行った。

練習終了後、本番を行う前に1対1モードか1対多モードで行うかを実験参加者に選んでもらった。

本番時のモードを選び終わった後、5分間のゴール誘導タスクを行ってもらった。本番ではロボットは予測違反モードで動作した。3つのゴールに順にロボットが到達した時点で実験終了とした。ゴールの基準はロボットの中心が数字を囲んでいる四角の中を通過したり停止したりすることとした。

本番終了後、動きの予測違反を行った際のロボットの印象をアンケートで測定し、実験を終了した。

2.4.2 実験 2

実験 1 においては本番時に1対1モードか1対多モードで動作するかを実験参加者が決定していたため、ロボットから受ける印象に抽象粒度の選好が影響を与える可能性がある。そこで、実験 2 では実験者が1対1モードか1対多モードの選択を行った。

3. 実験結果

3.1 関係の複雑性

アンケート (表 1) に対して一元配置分散分析を行った。

まずロボットのモードを実験参加者が選んだ場合 (実験 1) の結果を記す。実験の結果、表 1 より、「命令を拒否していると感じた」の評定値の平均は1対1水準、1対多水準それぞれ1.67, 3.50であり統計的に有意な差が確認された ($F(1,8) = 6.24, p < 0.05$)。また「動きを予測できた」の評定値の平均は1対1水準、1対多水準それぞれ5.33, 2.25であり統計的に有意な差が確認された

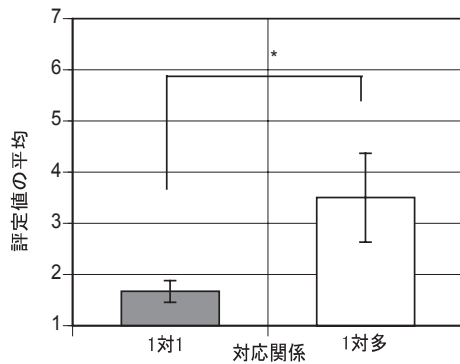


図 6 命令を拒否していると感じた。エラーバーは標準誤差。
 * $p < .05$

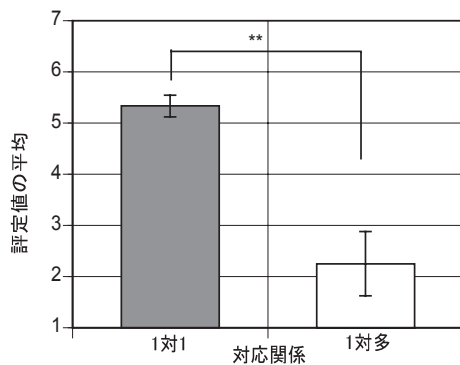


図 7 動きを予測できた。エラーバーは標準誤差。
 ** $p < .01$

($F(1, 8) = 0.00, p < 0.01$). その他のアンケートの項目からは有意な差は確認されなかった(付録の図 A.1 参照). このことから仮説にあった通り, 対応関係の単純なマッピングと複雑なマッピングとの間に受ける印象について差異が見られた.

次に, モード選択を実験者が行った場合(実験 2)の結果を示す. いずれのアンケート項目についても評定値の平均値に有意な差が見られなかった(付録の図 A.2 参照).

4. 議論

4.1 入出力関係の複雑さと意図帰属

実験参加者自身がゴール誘導タスク時のロボットの入出力関係の複雑さを選択した場合, ゴール誘導タスク実行にロボットが予測違反行動を出力すると, 1対1モードで操作した実験参加者よりも1対多モードで操作した実験参加者の方がより「命令を拒否していると感じた」と感じることがわかった. また, 「動きが予測できた」については1対多モードで操作した実験参加者よりも1対1モードで操作した実験参加者の方が評定値の平均は高かった. この2つの結果から本番時に事前にロボットのモードを説明しているにもかかわらず, 1対1モードよりも1対多モードの方がロボットの動きが予測できず命令を拒否していると感じていたことがわかる.

ロボットが自分の出した命令と異なる動きをしたとき,

実験参加者は「なぜ自分の命令を聞かなかったのか」と疑問に感じ, 「ロボットに別の思惑があるのではないかと考えたと考えられる. その状態からタスクを達成するためには, ロボットの出力された様々な行動からロボットの内面にある見えない抽象的なルール(意図)を考えなくてはならない.

抽象度の低い単純な動作を行う物体については, 関係性が単純なため心的状態を推測するまでもなく理解が出来る. しかし, 抽象度の高い動作を行う物体については, 1つ1つの観察できる振る舞いに注目すると一貫性のない行動の集まりと考えがちであり, その物体が何をしたいのかが理解できない事が多い. しかしそれらの出力された振る舞いから, 観察不可能である内面的な「心的状態(意図)」を推測することで, 一見何の繋がりのない様々な行動が一貫性を持ち理解が出来るようになりタスクをこなすことが出来ると考えられる(本研究では「振る舞い」は行動出力, 「心的状態」はゴールに向かうという「意図」).

このことからロボットの振る舞いの抽象度に関して「複雑でない単純な入出力関係」よりも「抽象的で複雑である入出力関係」である場合に, 参加者がそのロボットに対し「心的状態(意図)」を考え推測し, 意図性を感じたと考えられる.

4.2 ロボットの振る舞いの選好について

実験 1 と実験 2 の違いは実験参加者自らがゴール誘導タスク時のロボットの入出力関係の複雑さを選択したかどうかである. 実験 1 において実験参加者がロボットの入出力関係の複雑さを選択した場合, 入出力関係が複雑な場合に単純な場合よりも実験参加者がロボットの動作に対して意図を帰属した可能性があったが, 実験 2 において実験参加者が入出力関係の複雑さを選択した場合, アンケートのすべての項目で有意差が見られなかった. このことからロボットの振る舞いから意図性を感じたり心的状態を推測するには, ロボットの振る舞いに対する「好み」や「愛着」, 「興味」などの選好が関係している可能性があると考えられる.

このことから振る舞いの抽象化粒度の選好が心的状態の推測, 帰属に関係している可能性があると考えられる.

5. まとめ

本研究では, ロボットの振る舞いの抽象度に関して, 入出力マッピングの抽象化粒度を上げるとロボットに意図性を感じる可能性が高い事が示された. また, ロボットに意図性を感じるのにロボットの振る舞いに関する「好み」や「愛着」などの選好が関係している可能性を示唆した. この事から人とロボットとのインタラクションの中で, 抽象度の高い複雑な振る舞いを起こす事と振る舞いの抽象化粒度の選好がロボットから「意図性」を感じる上で必要である可能性があることが示された.

謝辞 本研究は MEXT 科研費 25119502 の助成を受けたものである。記して感謝する。

参考文献

- [1] アンドリュー・ホワイトウン, リチャード・バーン (編) : マキャベリの知性と心の理論の進化論 II, ナカニシヤ出版 (2004).
- [2] Dunbar, R.: *Grooming, Gossip, and the Evolution of Language*, Harvard University Press (1998).
- [3] 伊藤 昭, 寺田和憲: 人工知能と心の理論-心を持つロボットへの試み, 臨床発達心理学実践研究, Vol. 9, pp. 16-20 (2014).
- [4] 綾屋紗月, 熊谷晋一郎: 発達障害当事者研究, 医学書院 (2008).
- [5] Dennett, D. C.: *The Intentional Stance*, Cambridge, Mass, Bradford Books/MIT Press (1987).
- [6] Gergely, G., Nádasdy, Z., Csibra, G. and Bíró, S.: Taking the intentional stance at 12 months of age, *Cognition*, Vol. 56, No. 2, pp. 165-193 (1995).
- [7] Gergely, G., Bekkering, H. and Király, I.: Rational imitation in preverbal infants., *Nature*, Vol. 415, No. 6873, p. 755 (online), DOI: 10.1038/415755a (2002).

付 録

A.1 3輪オムニホイール型ロボットの運動学

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{pmatrix} = \frac{R^2 r}{\det A} \begin{pmatrix} -(\sqrt{3} \sin \theta) & \frac{1}{2}(\sqrt{3} \sin \theta) - 3 \cos \theta & \frac{1}{2}(\sqrt{3} \sin \theta) + 3 \cos \theta \\ (\sqrt{3} \cos \theta) & -\frac{1}{2}(\sqrt{3} \cos \theta) - 3 \sin \theta & -\frac{1}{2}(\sqrt{3} \cos \theta) - 3 \sin \theta \\ 1 & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \quad (\text{A.1})$$

$$\det A = \frac{3\sqrt{3}R}{2} \quad (\text{A.2})$$

A.2 分析結果

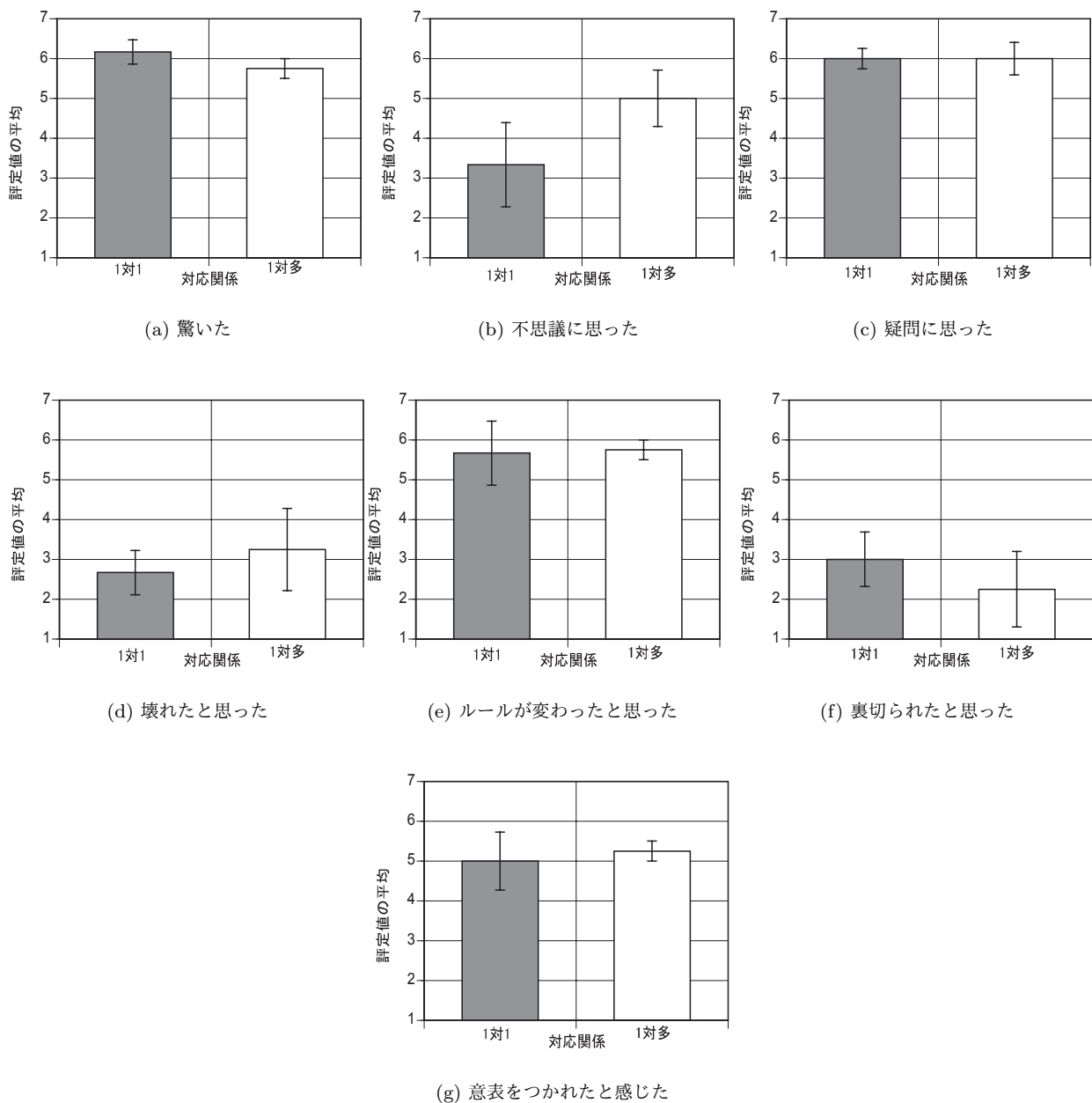


図 A.1 実験1のアンケート結果

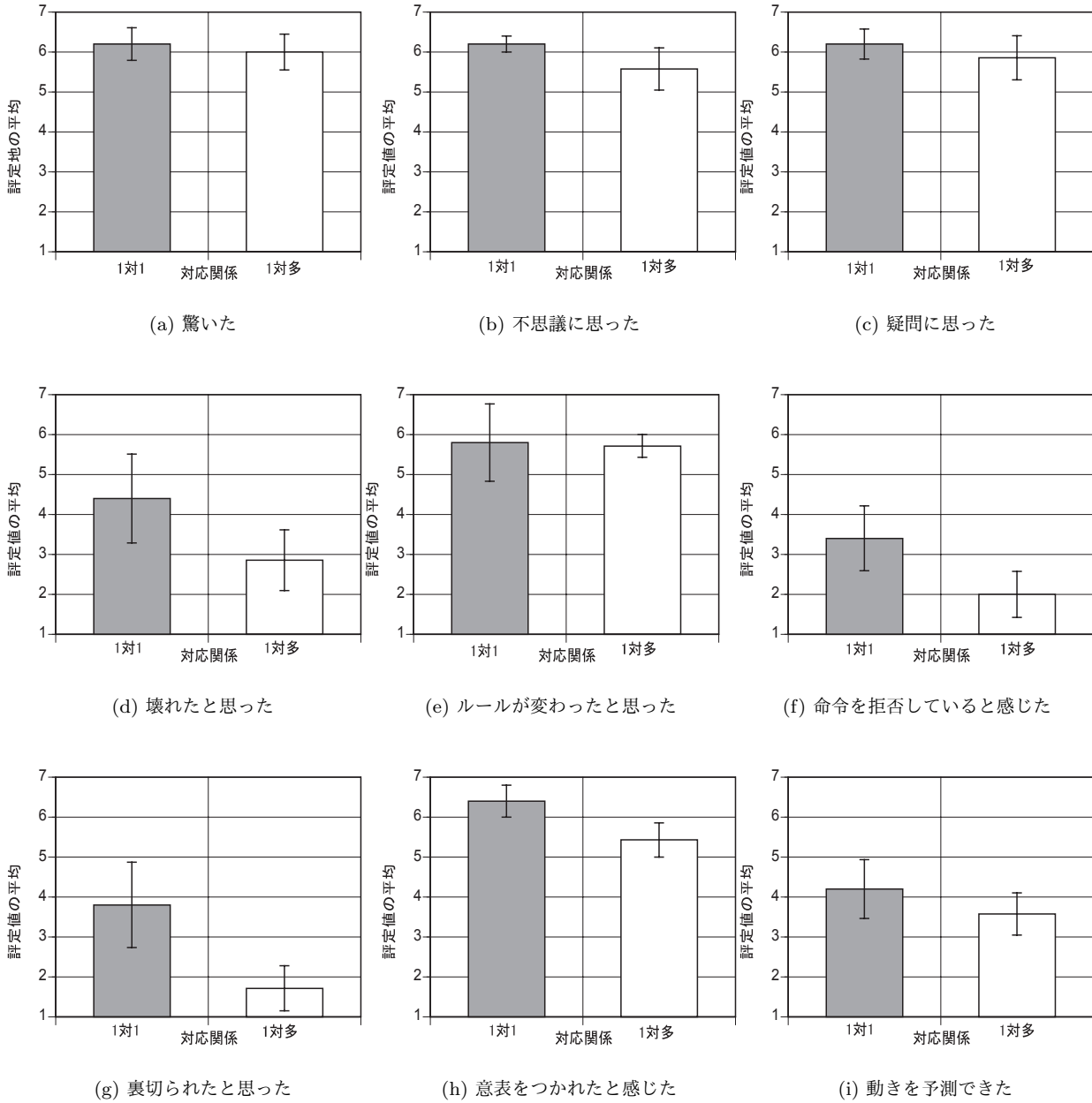


図 A-2 実験2のアンケート結果