ボーナス付きマッチングペニーゲームにおける人間からエージェントへの適応プロセスの解明

Experimental Investigation of Human Adaptation to Change in Agent Strategy in Competitive Two-Player Game

寺田 和憲[1*] 山田 誠二[2] 伊藤 昭[1]
Kazunori Terada[1] Seiji Yamada[2] Akira Ito[1]

[1] 岐阜大学 工学部 応用情報学科
[1] Gifu University, Faculty of Engineering, Department of Information Science
[2] 国立情報学研究所／総合研究大学院大学／東京工業大学
[2] National Institute of Informatics, SOKENDAI, Tokyo Institute of Technology

Abstract: We conducted an experimental investigation on human adaptation to change in an agent's strategy through a competitive two-player game. Modeling the process of human adaptation to agents is important for designing intelligent interface agents and adaptive user interfaces that learn a user's preferences and behavior strategy. However, few studies on human adaptation to such an agent have been done. We propose a human adaptation model for a two-player game. We prepared an on-line experimental system in which a participant and an agent play a repeated penny-matching game with a bonus round. We then conducted experiments in which different opponent agents (human or robot) change their strategy during the game. The experimental results indicated that, as expected, there is an adaptation phase when a human is confronted with a change in the opponent agent's strategy, and adaptation is faster when a human is competing with robot than with another human.

1 はじめに

人間とエージェント，ロボット間のインタラクションデザインを統合的に扱う HAI ヒューマンエージェントインタラクション [8] が活発に研究されている．そこでは，人間がエージェントと対峙したときに，そのエージェントのモデルを構成し，その行動を理解・予測する心理的プロセスを解明することが重要な課題である [9]．この解明により，人間によるエージェントのモデル化とその利用を踏まえた上で，エージェントの持つべきアピアランスや行動戦略を決定すること [7, 3] が可能となる．

さらに，我々は，この人間によるエージェントのモデル化とその利用における最も重要な課題の一つは，エージェントの行動戦略がオンラインで変化した場合に，それに対して人間がどのように適応していくかという「人間からエージェントへの適応プロセス」の解明であると考える．この適応プロセスを解明することで，人間が理解しやすく，適応しやすいエージェントや対話的システムの設計 [4] に指針を与えることが期待できる．

以上の背景から，本研究では，人間とエージェントが行うシンプルな二人対戦ゲームにおいて，ゲーム途中でエージェントの行動戦略が変化したときに，人間はその変化に適応できるのか，あるいはその適応にはどのような特性があるのかを実験的に解明することを目指す．この実験では，独立変数として 2 つの異なるエージェントを設定し，それらの差異が人間からエージェントへの適応に与える影響に注目する．

2 人間からエージェントへの適応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，お互いの利害が対立しており，相手の行動が観測可能で，その観測された行動を基に相手モデルを推定できる**競合ゲーム**（competitive game）を用いる．

*連絡先：岐阜大学工学部応用情報学科
〒 501-1193 岐阜県岐阜市柳戸 1-1
E-mail: terada@info.gifu-u.ac.jp

2.1 適応プロセスのモデル

我々は，競合ゲームにおける人間の適応プロセスを**適応フェーズ**と**利用フェーズ**の二つのフェーズによってモデル化する．二つのフェーズにおけるパフォーマンスを簡略化してグラフに表したものが図 1 である．通常，競合ゲームにおいて人間のプレーヤーは，相手のとった行動からその行動戦略モデルを同定し，それに応じて自分自身の行動戦略を決定する．しかし，観測できるのは表面的な振舞いである対戦相手の行為系列であり，相手の行動戦略モデルを直接知ることはできない．そのために，観測された振舞いから戦略モデルを推測する必要がある．しかし，異なる行動戦略が同一の行為系列を出力することがあるため，観測結果から行動戦略を同定することは不良設定問題である．ある場合には，振舞いの変化は相手を陥れるための揺さぶりのようなメタ戦略かもしれない．そのために即応的に相手の振舞いの変化に追従することは必ずしも得策ではない．

したがって，相手の行動戦略変化の兆候が観測された場合に，振舞いから推測される戦略を相手の現在の戦略として採用してよいか否かを見極めるためのフェーズが必要である．このフェーズでは，人間は自分自身の戦略を固定するのではなく，機械学習における**探査**（exploration）と**利用**（exploitation）を併せて行って，対戦相手エージェントの戦略変化に適応を試みると考えられる．そこで，このフェーズを**適応フェーズ**と呼ぶ．この適応フェーズでは，探査と利用が併用されるため，即時に適応が完了するのではなく，ある程度の時間幅を持ってパフォーマンスは漸次増加していくと考えられる（図 1）．

そして，一度適応が収束して相手の行動戦略モデルが固定されると，そのモデルに基づいて決まった自身の行動戦略に従って手を出し続ける**利用フェーズ**に入る．続いて，また相手の行動戦略が変化した兆候が見られれば，適応フェーズになるということを繰り返す．利用フェーズでは相手に勝ち続けるため，自身のパフォーマンスは高得点でほぼ一定に保たれるが，適応フェーズではその始まりで相手の戦略変化に対応できずにパフォーマンスが急激に低下して，その後は探査と利用を行いながら徐々にパフォーマンスを向上させていくと考えられる（図 1）．ここで，適応フェーズでモデルを更新するまでに必要なラウンド数 τ（図 1）の逆数 $\frac{1}{\tau}$ を**適応速度** (adaptation speed) と呼ぶ．

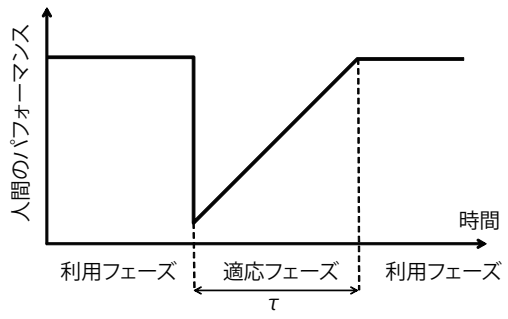

図 1: 利用-適応フェーズのパフォーマンス

2.2 対象の違いによる適応の違い

エージェントの外見は，人間がエージェントの認知的能力と行動能力を推測する際の重要な情報となる．観察者が対象であるエージェントをモデル化する際に用いる基準として**意図** (intentional)，**設計** (design)，**物理** (physical) の 3 つのスタンスの存在がデネットにより指摘されている [2, 6]．物理スタンスとは，主体の物理的組成，性質，物理法則に基づいて振舞を予測する戦略である．設計スタンスとは，物理スタンスで想定される物理的組成などの細部を無視し，主体が設計意図に基づいて作られていることを前提として，設計どおりに振舞うと予測する戦略である．そして，意図スタンスとは，主体の振舞が意図，信念，願望などの心的状態に基づいて合理的に生成されているという前提のもとに，振舞の起源となる心的状態を帰属した上で振舞を予測する戦略である．エージェントは機械である反面，意図的存在として捉えられることがある．意図スタンス採用のトリガーとして外見や自発性など様々な要因が考えられているが，寺田らはスタンスが振舞い予測のための戦略であることから，意外性が意図スタンス採用に寄与することを論証し，実験的に検証している [5]．

本研究では，人間からエージェントへの適応プロセスを明らかにするとともに，そのプロセスが機械と人間に対して異なるか否かを明らかにする．我々は，人間のプレーヤーは，機械の対戦相手に対しては単純で決定論的なモデルを構成し，人間の対戦相手に対しては複雑で確率的なモデルを構成すると考える．その結果，機械の対戦相手に対する適応は，人間のそれに対する適応よりも速いと予測される．このことは，次節において仮定として示され，適応速度を実験的に比較することで検証される．

3 2 人対戦ゲームによる参加者実験

本実験の目的は，人間からエージェントの適応プロセスに関して，次の 2 つの仮説を検証することである．

H1 相手の戦略変化に追従するための適応フェーズが存在する

H2 対ロボットの適応速度は対人間の場合よりも速い

本実験ではこれらの仮説の検証のために，相手の行動戦略を推定することが必要な競合ゲームとして，マッチングペニーゲームを改変した**ボーナス付きマッチングペニーゲーム**を用いる．

3.1 ボーナス付きマッチングペニーゲーム

マッチングペニーゲームは，2 名のプレーヤーが同時に硬貨を出し合い，表裏が一致していれば一方が得点（2 枚の硬貨）を貰え，不一致の場合はもう一方が得点を貰えるゼロ和ゲームである．このゲームはじゃんけんと同様にナッシュ均衡を満たす純粋戦略は存在せず，ランダムに（裏表を等確率で）出す混合戦略が唯一のナッシュ均衡であることが知られている．

本研究ではこのゲームの配点を改変し，6 回に 1 回の割合で周期的にボーナスラウンドを設ける．ボーナスラウンドでは，通常の 1 ラウンド得点 1 点の 20 倍である 20 点が得られるようにした．この改変によっても，ナッシュ均衡は毎回独立に裏表を等確率で出すことに変わりはないが，履歴を意識させる効果がある．

3.2 ボーナス付きマッチングペニーゲームにおける行動戦略

ボーナスラウンドの導入により，プレーヤは相手を陥れる戦略を採ることができる．例えば，通常得点期間である 1〜5 ラウンドの間に常に表を出すなどの規則的な手を出すことにより，相手にその規則を推定させ，ボーナスラウンドでも同じ規則で出すと思わせておいて，実際には逆を出すという**裏切り戦略**が生成可能になる．表 1 に，本実験で用いた，エージェントの具体的な行動戦略を示す．図中で，●と○はコインの表，裏であり，裏切り戦略に対して裏切らない戦略を**規則踏襲戦略**と呼ぶ．また，この 2 戦略の双方に「一様」と「交互」の 2 パターンがある．それぞれのパターンにおいて，2 つの戦略の違いは第 6 ラウンドで裏切るか否かの違いのみであり，1 から 5 ラウンドまでは同じ手である．本実験では表 1 の一様 2 パターンと交互 2 パターンの計 4 パターンそれぞれに対し，表から始まる場合と裏から始まる場合の 2 つのバリエーションを考慮し，全部で 8 パターンを用いることにした．

本研究では，実際に人間の適応プロセス手続きに従ってエージェントモデルの更新が行われているかどうかを調べるために，相手エージェントが対戦の途中で戦略を切り替える場合について参加者実験を行う．具体的には，6 ラウンドを 1 ゲームとし，10 ゲーム行う．対戦相手エージェントは最初の 3 ゲームでは規則踏襲戦略で行動し，第 4 ゲーム以降はすべて裏切り戦略で行動する．これは，最初の 3 ゲームで規則踏襲戦略の手を出すことで，参加者に対戦相手が容易に搾取可能な相手だと思わせ，第 4 ゲームの戦略変更を参加者に印象付ける狙いがある．また，第 4 ゲーム以降の参加者の適応を観察するために，第 4 ゲーム以降は戦略変化を行わない．すなわち，参加者が戦略変化に即時に追従することができれば，5 から 10 ゲームをすべて勝つことが可能な設定になっている．

表 1: 各ラウンドの得点と対戦相手エージェントの二つの行動戦略

ラウンド		1	2	3	4	5	6
得点		1	1	1	1	1	20
規則踏襲戦略	一様	●	●	●	●	●	●
	交互	●	○	●	○	●	○
裏切り戦略	一様	●	●	●	●	●	○
	交互	●	○	●	○	●	●

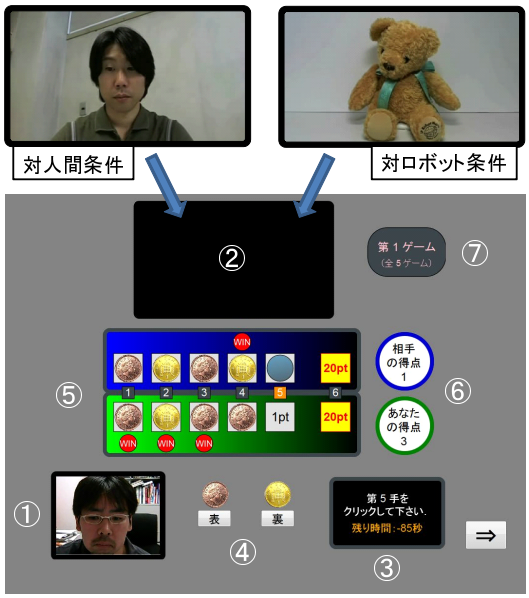

図 2: オンライン実験システムの画面

3.3 実験の設定と手続き

実験はすべて Web サーバーと接続した PC の Web ブラウザ（Firefox）上においてオンラインで行われた．Web サーバー上のプログラムは，JavaScript，ライブラリ Raphael[1]，jQuery webcam plugin[2]，そして HTML で開発された．

[1]http://raphaeljs.com/
[2]http://www.xarg.org/project/jquery-webcam-plugin/

ボーナス付きマッチングペニーゲームをプレイする画面を図 2 に示す．画面には，①ウェブカメラでキャプチャした参加者の顔，②対戦相手エージェント，③ゲーム進行の指示と参加者が一手を決定しなければならない制限時間（10 秒）の残り時間，④参加者が自分の次の手を決定するためのボタン，⑤現在までの両者の手と勝敗，各ラウンドのポイント数，⑥両者の得点，⑦現在のゲーム番号と全体のゲーム数，の情報が提示された．右下の「⇒」ボタンにより「ゲーム開始」と「次のゲームへ」への遷移が可能であった．

ウインドウ②に表示されたのは，図 2 の上にある人間とロボットの 2 種類のエージェントが動く動画であり，これらが 2 水準の独立変数となる．対人間条件の動画には，参加者と面識のない人間が，実際にこのゲームを対戦したときの様子を撮影したものを用いた．対ロボット条件では，手足と首が動くクマ型ロボット (IP RobotPhone) が手や頭を動かす 6 種類の動作をランダムに生成する様子を撮影したものを用いた．

実験の目的は,「オンラインゲームのユーザビリティ調査」であり，画面に映った相手と対戦するゲームであることが伝えられた．また，実験終了後に獲得したポイントに応じた容量のマイクロ SD カードが参加者に与えられることも伝えられた．

参加者は，練習のための 5 ゲームと，実験用の 10 ゲームを対戦することを求められた．練習用の 5 ゲームでは，6 ラウンド目のみ相手エージェントがどちらの手を出したかが分からないようになっており，相手の戦略推定を行わず，単にゲーム画面の操作に慣れてもらう目的がある．

各戦略の 4 パターンの出現順は，参加者ごとにことなる乱数によって決定した．ただし，同一パターンが連続して採られることはないようにした．参加者は，各水準に 14 名づつの計 28 名で，情報系の大学生と大学院生（平均年齢 21.3 歳，標準偏差 2.1 歳）であった．

3.4 観測

各ラウンドにおける参加者の勝敗を測定項目とした．本実験の目的は，参加者が対戦相手の戦略モデルをいかに構築し，利用するかを調べることあるが，参加者の獲得した相手モデルを直接調べることはできない．しかし，勝敗によって間接的に参加者が対戦相手の戦略モデルを推定していたか否かを知ることができる．戦略モデルを持たない場合，参加者の勝敗は等確率（0.5）であるが，正しい戦略モデルをもてば参加者が勝つ確率は等確率より高くなり，間違ったモデルであれば低くなる．

戦略予測については，特に第 6 ラウンドの結果に注目する．第 6 ラウンドの勝敗は，参加者が対戦相手がどの行動戦略（規則踏襲戦略か裏切り戦略）を採ると推定したかを表す．

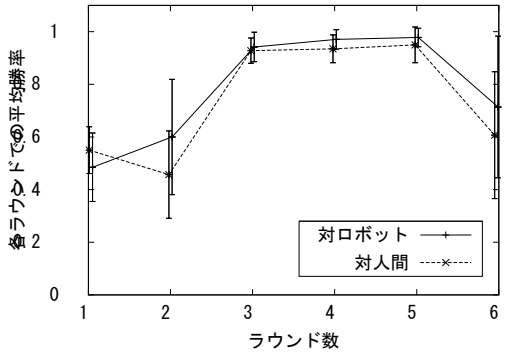

図 3: 各ラウンドでの平均勝率

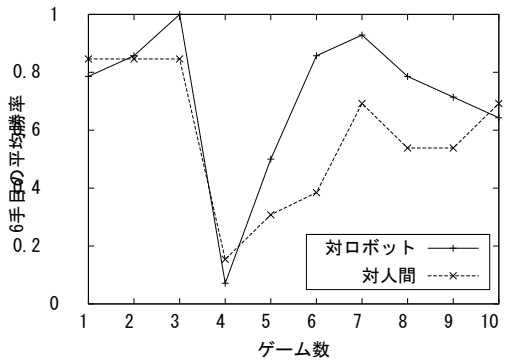

図 4: 第 6 ラウンドの平均勝率

3.5 実験結果

図 3 は各ラウンドにおいて勝った参加者の割合[3]（以下，勝率と呼ぶ）の全 10 ゲームの平均と標準偏差を条件ごとにグラフに表したものである．図より，両条件ともに，最初の 1，2 ラウンドでは勝率がチャンスレベルの 0.5 付近であるのに対し，3 ラウンドから 5 ラウンドでは 0.9 以上に上昇していることが分かる．これは，参加者が 1，2 ラウンドのうちにエージェントの出す手の規則性を理解し，3 ラウンド以降に適応したことを意味する．しかし，第 6 ラウンドでは，エージェントが規則踏襲戦略と裏切り戦略を使い分けているために，勝率が下がっている．

図 4 は，第 6 ラウンドの勝率のゲームごとの推移を表している．図より，ゲーム間で勝率が大きく異なることがわかる．図 3 で第 6 ラウンドの標準偏差が大きいのはこの変動のためである．第 4 ゲームではほとん

[3]条件ごとに計算するために分母は常に 14 である．

表 2: 第 3 ゲームの勝率との違いについての χ^2 検定の結果

	ゲーム数						
	4	5	6	7	8	9	10
対ロボット	**	**				*	*
対人間	**	**	*				

* $p < .05$, ** $p < .01$

どの参加者が負けているが，これは対戦相手エージェントが第 4 ゲームで戦略を規則踏襲から裏切りに変更したためである．第 4 ゲーム以降，勝率は両条件ともに回復している．

最初の 3 ゲームでは勝率が安定していることから，参加者は規則踏襲戦略に適応しており，この期間は利用フェーズであると考えられる．参加者は第 4 ゲームの 6 ラウンド目の相手の手によって相手エージェントが戦略を変化させたことに気づいたと考えられる．そこで，適応フェーズの開始を第 4 ゲームとする．適応フェーズは，第 3 ゲームの勝率を基準として用い，勝率の差が観測されなくなった時に終了と判定できる．

そこで，第 3 ゲームの勝率と第 4 ゲーム以降の勝率に有意差があるかどうかをカイ二乗検定によって検定した．検定結果を表 2 に示す．表より，両条件ともに勝率が回復するまでに少なくとも 2 ゲームを要していることが分かる．勝率が即時に回復するのではなく，時間とともに漸次増加していることは適応フェーズの存在を示唆しており，3 章の **H1** は支持されたと言える．また，適応フェーズ内で実際に探査と利用の併用が行われたかどうかについては，次章で考察する．

一方，勝率の回復は条件間で異なっている（表 2）．対ロボット条件では 6 ゲーム目で，対人間条件では 7 ゲーム目と，対人間条件では対ロボット条件よりも 1 ゲーム多く必要であったことが分かる．なお，2.1 章の適応速度は，対ロボットで 0.50，対人間で 0.33 となる．この結果から対戦相手がロボットの場合には人間の場合よりも適応が速いことが示唆され，**H2** が支持されたと考えられる．

4 考察

4.1 2 条件での適応フェーズの違い

図 4 の勝率の変化は参加者を群として見たときの傾向を表している．そのため，適応フェーズである第 4 ゲームから第 7 ゲームにかけての傾きは間接的には，2.1 章の適応速度を表しているが，参加者個人のパフォーマンスの変化を直接表しているわけではない．また，2.1 章で我々がパフォーマンスの漸次増加の根拠として考えた，適応フェーズにおける探査と利用の併用が行われていることの直接の根拠とはならない．なぜなら，図 4 は勝った参加者の割合であるために，勝率の漸次増加の原因として，1) 参加者が戦略を 1 回だけ切り替えただけであるのにもかかわらず，参加者間で戦略変化のタイミングが分散していた可能性，2) 自身の戦略を変化させながら探索的な適応を行っていた可能性の双方が考えられるからである．

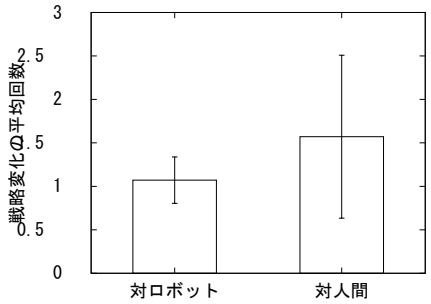

図 5: 5 から 7 ゲームで参加者が戦略を変化させた回数の平均

そこで，適応フェーズにおいて各参加者が実際に戦略を変化させた回数を調べた．適応フェーズである第 5 ゲームから第 7 ゲームにおいて，ひとつ前のゲームと異なる戦略を取ったゲームの数の平均と標準偏差を図 5 に示す．t 検定の結果，ロボット，人間条件間の戦略変化回数の平均値の差に有意傾向が確認された ($p = 0.066$)．5 から 7 ゲームの間に参加者が戦略を変化させる機会は 3 回あるが，対ロボットの場合は平均 1.07 回，対人間の場合は平均 1.57 回戦略を変化させている．もし，参加者がエージェントが途中で 1 回だけ戦略を変更することを知っていたのであれば，最適な適応は，第 4 ゲームでのエージェントの戦略変更を認識後，第 5 ゲームで即座に裏切り戦略に追従することである．しかし，現実にはエージェントの戦略を知ることは不可能であり，図 4 から分かるように，第 5 ゲームで戦略を変化させ即座に追従した参加者は対ロボットでも 5 割程度である．

対ロボットの参加者の第 5 ゲームから第 7 ゲームでの戦略変化の回数がほぼ 1 回のみであることは，図 4 の第 4 ゲームから第 7 ゲームにおいて，対ロボット条件での勝率が単調増加しているように見えるのが，個々の参加者の探索的な戦略変化の結果ではなく，参加者間で適応開始のタイミングが分散していたことに起因することを意味する．一方で，対人間の参加者の第 5 ゲームから第 7 ゲームでの戦略変化の回数が平均 1.57 回であることは，この期間に探索的な戦略変更（探査）を行っていたことを示唆する．

4.2 対戦相手の違い

エージェントの振舞いは全く同じであったのに，なぜアピアランスが人間だとロボットよりも適応速度が遅くなるのかについて考察する．競合状態における適切な振舞いが意図帰属に貢献することが知られている [5]．規則的な手は容易に看破され搾取されるので，6 ラウンド目の裏切りがなければ，最初の 5 手に規則的なパターンを出す意味はない．しかし，6 ラウンド目に勝つという目的のもとでは，相手に，自分は規則的な手を出すと信じさせて，裏切るのがボーナスポイントというゲームのルールを最大に活用した合理的戦略である．相手の信念を利用して利己的に振舞うためには心の理論 [1] が必要である．5 手の規則的なパターンがその後の裏切りに繋がる意図的な振舞いの前兆だと考えるのか，単なる設計的な振舞いだと考えるのかによって，6 手目の選択が異なると考えられる．

最初の 3 ゲームでは，両条件ともに対戦相手エージェントを設計スタンスでとらえていたと考えられる．しかし，4 ゲームでエージェントが戦略を変化させた際，対人間では 4 ゲーム目の戦略変化が意図的な振舞いとして感じられ，対ロボットでは単に別の規則として設計スタンスのままでとらえられたと考えられる．相手を意図スタンスでとらえると，相手の意図を推定する読みが発生する．また，相手に自分の戦略を読まれる可能性があるため，固定的戦略は不利になる．その結果，戦略が固定されず，結果として適応速度が遅くなったと解釈される．一方，対ロボットの場合は，4 ゲーム目の戦略変化を意図ではなくメタ規則としてとらえる傾向が強く，素直に相手の戦略変化に追従したために適応速度が速く，さらに 8 ゲーム目での周期的なメタ規則による戦略変化をより強く予測したために，図 4 と表 2 で 9 ゲーム以降に有意に勝率が下がったと考えられる．

5 まとめ

本研究では，人間とエージェントの利害が対立した状況を競合ゲームによって作り出し，そこでエージェントのオンラインでの行動戦略変化に人間がどのように適応するのかを調べた．まず，適応プロセスを利用フェーズと適応フェーズの二つのフェーズによってモデル化した．適応フェーズでは人間は人間は自分自身の戦略を固定するのではなく，探査と利用を併用し，相手エージェントの戦略変化に適応する．利用フェーズでは，適応フェーズで獲得した相手の行動戦略モデルに従って搾取的に行動する．さらに，適応速度に注目し，対人間の場合よりも対ロボットの場合の方が適応速度が速いという仮説を導出した．モデルの妥当性と仮説の検証のために，ボーナス付きマッチングペニーゲームを用いた参加者実験を行った．実験の結果，相手の戦略変化に対し，探査と利用を併用した適応フェーズが存在することが確認された．また，対人間の場合よりも対ロボットの場合の方が適応速度が速いことが確認された．

謝辞

本研究は，平成 23 年度国立情報学研究所共同研究（戦略研究公募型）『観察者の持つエージェントのモデルと実際の振舞いのギャップが解消される過程のモデル化』の支援を受けた．記して感謝いたします．

参考文献

[1] Simon Baron-Cohen. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, 1995.

[2] Daniel C. Dennett. *The Intentional Stance*. Cambridge, Mass, Bradford Books/MIT Press, 1987.

[3] 小松孝徳, 山田誠二．適応ギャップがユーザのエージェントに対する印象変化に 与える影響．人工知能学会誌, Vol. 24, No. 2, pp. 232–240, 2009.

[4] Talia Lavie and Joachim Meyer. Benefits and costs of adaptive user interfaces. *Journal International Journal of Human-Computer Studies*, Vol. 68, pp. 508–524, 2010.

[5] 寺田和憲, 伊藤昭. 人間はロボットに騙されるか？ -ロボットの意外な振舞は意図帰属の原因となる-. 日本ロボット学会誌, Vol. 29, No. 5, pp. 43–52, 2011.

[6] 寺田和憲, 岩瀬寛, 伊藤昭. Dennett の論考による 3 つのスタンスの検証. 電子情報通信学会論文誌 (A), Vol. J95-A, No. 1, 2012.

[7] A. Tinwell and M. Grimshaw. Bridging the uncanny: an impossible traverse? In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, MindTrek '09, pp. 66–73, 2009.

[8] 山田誠二（監著）. 人とロボットの＜間＞をデザインする. 東京電機大学出版局, 2007.

[9] 山田誠二. Hai 研究のオリジナリティ. 人工知能学会誌, Vol. 24, No. 6, pp. 810–817, 2009.