人の行動決定過程におけるメタ戦略の存在とその処理過程

Existence of "Meta-strategy" in human action decision and its processing

横山　絢美 [1*]　大森　隆司 [2]
Ayami Yokoyama[1]　Takashi Omori[2]

[1] 玉川大学大学院工学研究科
[1] Graduate School of Engineering, Tamagawa University
[2] 玉川大学工学部
[2] College of Engineering, Tamagawa University

Abstract: We use various types of strategy to realize a smooth interaction with others. We can estimate intention of others and determine own action according to the other's intention or can induce the others' intention and actions as we intended. In our past study, we have modeled the process of action decision based on intention estimation of others and evaluated its effectiveness by a computer simulation in cooperative situation. As a result, we presented necessity of "Meta-strategy" such as a choice of action strategy from possible ones to decide the action of next moment. In this study, we analyzed human action in cooperative game and we discuss on a feature of human "Meta-strategy" for action decision process in such an interactive situation.

1　はじめに

我々は日常生活において常に他者と関わりながら生活している．その中で，他者との協調は重要で，我々は初対面の相手であっても，自然と協調行動をとることができる．では，いかにして他者とのインタラクションを実現しているのだろうか．

我々は他者とのインタラクションの際，互いに相手の状況や行動を観察することで他者の意図やプランを推定し，それに合わせて自身の行動を決定する．あるいは，自己の目的を行動や言語，身振りなどで他者に示し，自己の意図に合わせてもらう事もある．つまり，我々は他者の意図推定に基づいた振舞いをしていると言えるだろう．

では，このような他者の意図推定に基づいたインタラクションの実現を可能にしている心的メカニズムとはどのようなものだろうか．

我々はこれまでに，他者意図の推定に基づく自己と他者の行動決定過程の計算モデルを構築し，計算機シミュレーションを用いてその妥当性を検証した．また，その検証から状況に応じて次に取りうる戦略を選択する「メタ戦略」の必要性を示してきた．しかし，シミュレーションだけでは，実際に我々の取っている戦略について議論する事はできない．

そこで，本研究では協調ゲームにおける人間の行動について解析する．例えば，人は戦略の異なる他者に対してどのように振舞うのか，あるいは，シミュレーションの際に想定した行動戦略と被験者の振舞いには，どの程度の一致性があるのか，さらには，我々が想定していた以外の振舞いは存在するのだろうかという事について，行動実験と計算機シミュレーションの結果を合わせて検証し，我々の持つ「メタ戦略」の処理過程について議論する．

2　意図推定に基づく行動決定

2.1　行動決定戦略

我々は他者とのインタラクションにおいて，他者の行動や発話，振舞い等の情報を観察し，「他者がこのような場面でこういった振る舞いをするのは，A という意図を持っているからだろう」というように考える [1]．つまり，自己の経験と照らし合わせて他者の意図を推定し，自身の行動を決定して円滑なコミュニケーションを図っていると考えられる．我々は，このような他者から自己への一方向の行動決定をする戦略を「受動的な行動決定」と呼ぶ．

*玉川大学大学院工学研究科
〒 194-8610 東京都町田市玉川学園 6-1-1
E-mail: ykyma6er@engs.tamagawa.ac.jp

一方で，この「受動的な行動決定」戦略では他者の意図に従った行動決定は可能であるが，他者の意図が必ずしも自分との望んだものであるとは限らない．状況によって，自身の意図を他者に伝え，自己の意図に合うように誘導するように振舞うこともあるだろう．我々は，このように自己の意図を明示的に他者に示して自己の意図を推定させ，他者の意図あるいは行動を誘導しようとする戦略を「能動的な行動決定」と呼ぶ．

2.2 意図推定モデル

他者と円滑なコミュニケーションを実現する仕組みはどのようなものなのだろうか．図 1 は「受動的な行動決定」戦略 (以下「受動的戦略」) の概念図である．

「受動的戦略」では，他者の意図が基準となる．しかし，他者の意図は外からは観察することの出来ない隠れ状態である．そこで，我々は意図が反映されやすい他者の行動を観察し，自身がこれまでの経験などから得た知識と照らし合わせて他者の意図を推定する．そして，この推定した意図を基に自己の意図を決定し，自己の意図に合った行動を取る．

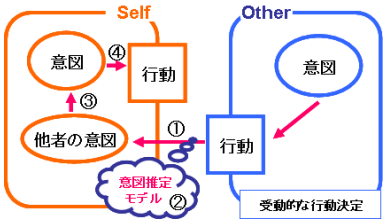

図 1: 受動的な行動決定モデル

一方,「能動的な行動決定」戦略 (以下,「能動的戦略」) の概念図を図 2 に示す．

「能動的戦略」では，自己の意図が規準となる．つまり，まず自己が達成したい意図 (目標) を決定し，それを他者に明示的に行動で示して，他者に自身の行動を観察させて自己の意図を推定させる．つまり，明示的な自己の働きかけが他者の意図に影響を与え，他者の状態や行動に変化を引き起こさせ，他者は自己にとって望ましい行動を行なう．

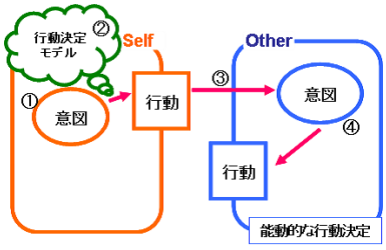

図 2: 能動的な行動決定モデル

2.3 推定のレベル

我々はこれまでに，日常生活で行なっている行動決定の戦略には,「受動的戦略」と「能動的戦略」があると述べてきた．

「受動的戦略」とは，他者から自己への一方向的な行動決定である．そこで，我々がまず思い描く振舞いは，他者の行動からその意図を推定し，それにもとづいて自身の行動を決定するといった状況だろう．我々はこのような行動決定の戦略を「Lv.1 の戦略」と呼ぶ．

しかし,「受動的戦略」は，単に他者の意図を推定する以外にも,「自身は他者からどのように思われているのだろう」というように，一段深く推定する戦略もありえる．我々はこのような行動決定の戦略を「Lv.2 の戦略」と呼ぶ．

一方,「能動的戦略」とは，まず自己の意図を決定する．すると，他者に目指して欲しい意図は自然と決まる．そこで，自身は自己の意図を明示的に他者に伝える行動を見せて他者の意図に影響を与え，結果的に他者の意図を誘導しようという戦略である．我々はこのような行動決定の戦略を「Lv.0*の戦略」と呼ぶ．

また,「受動的戦略」,「能動的戦略」のどちらでもないが，我々の振舞いとして，他者の意図推定など行なわず，自己の目標に向かって突き進むという戦略もありえる．我々はこのような行動決定の戦略を「Lv.0 の戦略」と呼ぶ．

2.4 人間のメタ戦略

我々は,「受動的戦略」あるいは「能動的戦略」といったいくつかの戦略を持っており，これらは「推定のレベル」としてより細かく分類することができる [2]．このことから，我々はその時々の状況に応じて，いくつかの戦略を切り替えることで，他者との円滑なコミュニケーションを実現しているのではないかと考えられる．

しかし，もしそうならば常に変動する社会において，この戦略をどのように選んで他者との円滑なインタラクションを実現しているのだろうか．我々は，これまでに各戦略の計算モデルを用いた協調課題を計算機シミュレーションで行なってきた．そして，この結果から円滑なインタラクションの実現のためには，状況に応じて戦略を選択する必要性，つまり「メタ戦略」の存在が示唆された．

しかし，これらはあくまでシミュレーションの結果であり，実際我々がそのような戦略を取っているかという事については定かではない．そこで，本研究では同様の協調課題を「人間 vs シミュレーション」で実現し，実験中の人間の振舞いや思考過程は，計算機シミュレーションで想定した行動戦略とどの程度一致性があるのか．また，人は戦略の異なる他者に対してどのよ

うに振舞うのか．更には，我々が想定していた戦略以外の振舞いも存在するのだろうか．といった事について解析する．そしてこれらの結果から導かれる人間の「メタ戦略」の処理過程について議論する．

3　モデルの評価実験 (Hunter Task)

3.1　実験環境とタスク

提案モデルの動作確認のため，我々は自己と他者の協調ゲーム (Hunter Task) を題材に，各戦略の振舞いや組合せ，得意/不得意な場面について検証するため，シミュレーションを行なった．

Hunter Task とは，2 体のハンターが互いに異なる獲物を捕獲する課題である．課題では 20 × 20 のトーラス状のグリッドワールドに，ハンター 2 体と獲物 2 体が存在する．この課題では，それぞれのハンターにとって，もう一方のハンターが他者となる．

ハンターと獲物は各時刻に 1 マスずつ行動し，獲物は上 20%，右 40%，停止 40%で確率的に行動する．そして，各々の戦略を持ったハンターが互いに異なる獲物を捕獲するとタスク解決となる．また，このタスクを効率的に解決するためには,「他者はどの獲物を狙っているだろう.」というように，他者の意図を推定しながらタスクに取り組む必要がある [3].

また，シミュレーションを行なうにあたり，ハンター 1 体と獲物 1 体で，意図推定や協調する必要がない課題を事前に行なった．この事前学習により，各ハンターは獲物を獲得するために必要な,「ある場所 s にいる時，行動 a を取れば自己の狙う獲物を捕獲することが出来る」という知識を強化学習 (Q 学習) によって，あらかじめ獲得させている．

3.2　意図の変数化

シミュレーションを行なうにあたり，ハンターの行動決定過程を定式化する必要がある．本研究では，ハンターの状態を s (state)，行動を a(action)，意図 (目標) を G(goal) として，この 3 変数を組み合わせて行動決定関数を定義する．つまり，これら 3 変数の同時生起確率分布は $P(a, s, G)$ と表す．

例えば，ハンターの意図 G は，現在の状態 s と行動 a から $P(G|a, s)$ として表すことができる．ハンターの行動 a についても同様で，現在の状態 s と意図 G が観察出来るならば $P(a \mid s, G)$ として表現することが出来る．さらに，ハンターの状態 s も，現在取っている行動 a と意図 G により，$P(s|a, G)$ と表現できる．

そして，これら 3 つの行動決定関数を応用すると，各戦略の処理過程も次のように表現できる．

Lv.1 の戦略は，まず，自己は他者の行動を観察する．そして，現在の他者の状態 s_o と行動 a_o を得て，自己の行動決定関数 (1) 式に代入し，他者の意図を推定する．そこで，自己はこの推定した他者の意図に合わせて自身の意図を決定し，現在の状態 s_s において自己の意図 G_s を達成するための取るべき行動を (2) 式によって決定する．

$$\tilde{G}_o = \underset{G}{\operatorname{argmax}} P(G|s_o, a_o) \tag{1}$$

$$a_s = \underset{a}{\operatorname{argmax}} P(a|s_s, G_s) \tag{2}$$

Lv.2 の戦略は，自己は現在の状態 s_s と行動 a_s を自身の行動決定関数 (3) 式に入力し,「他者が推定しているだろう自己の意図」を推定する．そこで，自己は「他者が推定しているだろう自己の意図」から他者の意図を推定し，それに合わせて自己の意図を決定する．そして，自己の意図を達成するために取るべき行動を (4) 式によって決定する．

$$\tilde{G}_s = \underset{G}{\operatorname{argmax}} P(G|s_s, a_s) \tag{3}$$

$$a_s = \underset{a}{\operatorname{argmax}} P(a|s_s, G_s) \tag{4}$$

Lv.0 の戦略は，自己は意図を達成するために取るべき行動を (5) 式によって決定し，この結果にのみ従って行動する．

$$a_s = \underset{a}{\operatorname{argmax}} P(a|s_s, G_s) \tag{5}$$

Lv.0*の戦略は，自己は，まず達成したい意図を決定する．すると他者に目指して欲しい目標 (意図) は必然的に決定する．Hunter Task では，他者には自己が決定した意図とは異なる意図によって行動して欲しいと考える．そこで，自己は意図を明示的に他者に伝えるために取るべき行動を (6) 式によって決定する．本タスクにおいて，ハンターは強化学習によってあらかじめ環境に対する知識を持っていることから，ここでの明示的な行動とは，ハンター (自己) が狙っている獲物とそうでない獲物を狙っている各々の場合の確率の差が最大となるような行動になる．

$$a_s = \underset{a}{\operatorname{argmax}}(P(G_s|s_s, a_s) - P(\tilde{G}_o|s_s, a_s)) \tag{6}$$

3.3　タスク解決までの効率性

シミュレーションでは，ハンターの取る各戦略の組合せを検討した．図 3 は，シミュレーションにおいて，各戦略の組合せ間でタスク解決までに要したステップ数 (10000 試行平均) の結果を示している．

ここではっきりと分かっている事は，ハンター 2 体のうち少なくとも一方は Lv.1 の戦略を取る必要があるという事である．つまり，どちらかが他者に合わせない限り上手くタスクを解決することはできない．

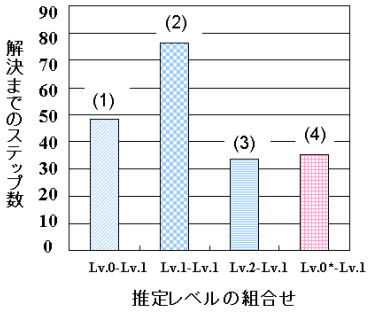

図 3: タスク解決までの効率性

図 3 の (1)〜(3) は，ハンターが互いに受動的戦略である場合のステップ数である．この中で，(1)「Lv.0 ― Lv.1」，(3) 「Lv.1 ― Lv.2」の組合せは少ないステップ数でタスクを効率的に解決しているのに対し，(2)「Lv.1 ― Lv.1」の組合せは解決までのステップ数が多い．これは，各ハンターの振舞いの中で，互いに意図の読み合いや目標修正が頻発したことによる．

一方，これらのデータと能動的戦略 (4)「Lv.0*― Lv.1」の組合せを比較すると,「受動的戦略」で最もステップ数の少ない (3)「Lv.2 ― Lv.1」の組合せとほぼ同程度のステップ数で解決している．そこで，我々はこの 2 つの事例についてより細かく評価した．

3.4 各戦略の適応範囲

ハンターと獲物のランダムな初期配置を 100 種類用意し，同様のシミュレーションを行なった．その結果,「受動的戦略」と「能動的戦略」の組合せで，解決までの要ステップ数の差が特に大きい事例の存在が見えてきた (図 4 ○部)．このことから,「受動的戦略」，あるいは「能動的戦略」には，それぞれが得意 / 不得意な場面が存在するのではないかと推測できる．

そこで，我々はこのような事例における初期配置についてより詳しく解析した．その結果，図 5 に示すような初期配置に依存するいくつかの傾向が見えてきた．

図 5(左) に示すように，例えば獲物がハンターを挟み，対称の位置で行動しているような場面では,「能動的戦略」によって明示的な行動を取るよりも「受動的戦略」を取った方が有効である (図 4 赤○)．

しかし，図 5(中央) に示すように，例えば獲物が自己より離れた位置で，比較的近い配置で行動しているよ

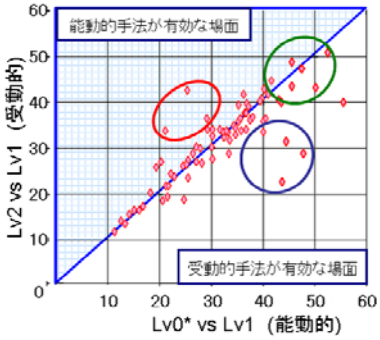

図 4: 受動的/能動的戦略の効果

図 5: モデルの適応範囲

うな場面では,「能動的戦略」の効果が発揮される (図 4 青○)．

さらに，図 5(右) に示すように，獲物が同位置、あるいは非常に近い配置で行動しているような場面には,「受動的戦略」,「能動的戦略」のどちらを取っても，タスクの効率的な解決は難しい (図 4 緑○)．

3.5 メタ戦略の重要性

我々は当初，円滑なインタラクション実現のためには，どのような戦略が最適であるのかと考えて Hunter Task を題材とした研究を行なった．しかし，解析の結果，最適戦略が存在するのではなく，その時々の状況に応じて戦略を選択する必要があるという事が分かってきた．つまり，我々は現在の状況を判断し，適切な行動を選択するため，より上位の「メタ戦略」に従って行動決定している．

このシミュレーション結果から，インタラクション過程についての仮説がいくつか見えてきた．では，実際に我々はどのようにしてこの「メタ戦略」を実現しているのだろうか．

しかし，これまでのシミュレーション結果だけでは，人間がどのような戦略を取っているかについては定かでない．そこで，我々は Hunter Task における人間の行動について解析し，シミュレーション結果と照らし合わせることで，我々の持つ「メタ戦略」の処理過程をより詳しく見る事にする．

4　行動実験

4.1　実験環境とタスク

被験者は大学生，大学院生合わせて 10 名で，実験課題は計算機シミュレーションと同様の Hunter Task を行なってもらった．

Hunter Task において，獲物の行動はシミュレーションと同様で，ハンターと獲物の初期配置は，シミュレーションの結果「受動的な戦略」が特に有効であった配置と,「能動的な戦略」が有効であった配置，更にはどちらの戦略を取っても解決が難しかった配置をそれぞれ４パターン計 12 種類の初期配置を用意した．その状況で，ハンター (他者) はそれぞれ Lv.0，Lv.1，Lv.2，Lv.0*の戦略を取る．

被験者には，相手ハンターと協力して別々の獲物を捕獲するよう教示し，ハンター (自己) を方向キーで操作してもらった．また，実験中は常に自己の狙っている獲物や実験場面，ハンター (他者) について感じたことや自己が振舞っている際に思ったことなどは，全て口頭で報告するよう求めた．

また，実験中の音声を記録するため，被験者にはイヤホンマイクを装着してもらい，実験中の被験者の行動は HyperCam2 を用いて Windows 画面上の動きをキャプチャーし，ムービーファイルとして保存した．

4.2　評価

行動実験中のハンター及び獲物の行動履歴から，シミュレーションで用いた Q 値を利用し，被験者がどの戦略に従って行動しているかについて評価した．

被験者の行動履歴 (各時刻における状態 s と行動 a) から，被験者が Lv.0，Lv.1，Lv.2，Lv.0*それぞれの戦略に従って行動していると仮定した際の Q 値を求め，この４つの Q 値を Soft-max を用いて確率値に変更する．この一つ一つの確率値から，各瞬間に被験者がどの戦略を取っていたかを推定する事は可能であるが，この推定値を何によって評価すれば良いかについては議論する必要がある．

そこで，我々は各試行における戦略の変化ではなく，被験者が特定の戦略に従って行動するハンター (他者) に対して取る戦略の傾向を見ることにした．

我々は，求めた４つの Q 値を変換した確率値の各試行中の平均を求め，その中の最大値を”１”それ以外を”０”と割り当てた．そして，各試行中のステップ数分のデータを平均し，被験者の戦略傾向を評価した．その結果を図 6〜8 に示す．

横軸に行動実験においてハンター (他者) が取った行動戦略，縦軸に各戦略を取った確率を示しており，各グラフにおけるデータは，それぞれ被験者 10 人の平均値である．

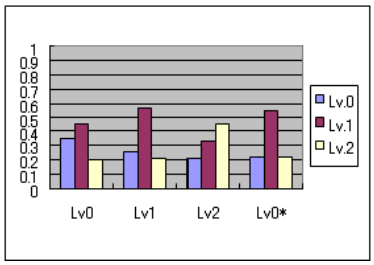

図 6: 被験者の行動戦略 (能動的戦略が有効な初期配置)

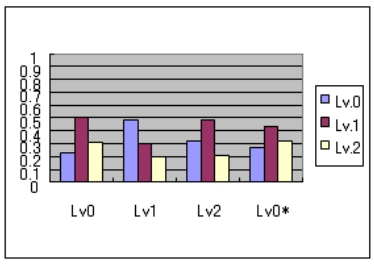

図 7: 被験者の行動戦略 (受動的戦略が有効な初期配置)

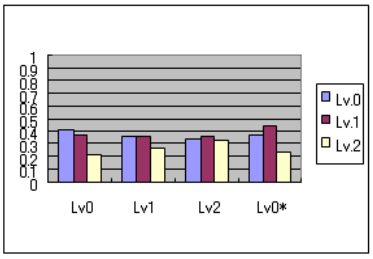

図 8: 被験者の行動戦略 (どの戦略でも難しい初期配置)

図 6 はシミュレーションにおいて能動的な戦略の効果が顕著に見られた初期配置に対し，被験者がどのレベルで行動していたかを示している．この初期配置において，コンピュータが Lv.0*，Lv.1 の戦略であれば，被験者は Lv.0*と相性の良い Lv.1 の戦略を取る確率が高く，この 2 つの戦略については効率よくタスクを解決できていることが分かる．

また，この初期配置においては，コンピュータの戦略が Lv.1，Lv.2 の場合，被験者はそれぞれ Lv.1 の戦略，Lv.2 の戦略を取る確率が高い．しかし，この状況では Lv.1 の戦略に対しては Lv.0 又は Lv.2 の戦略，Lv.2 の戦略に対しては Lv.1 の戦略が最も効率的な戦略なのだが，なぜか人間はこれらの行動を取っていない．

一方，図 7 は受動的な戦略の効果が顕著に見られた初期配置に対する被験者の振舞いを示している．このような初期配置では，コンピュータが Lv.0，Lv.1，Lv.2 の戦略を取ると被験者はコンピュータに合わせ，それぞれの戦略と相性の良い Lv.1，Lv.0，Lv.1 の戦略を取っている．

また，このデータではその差が明確にはなっていないが，Lv.0*についても，最適戦略である Lv.1 の戦略を取る確率は高い．つまり，被験者は他者の行動や意図を推定し，それに合わせて自己の振舞いを変更することで，効率的にタスクを解決していることが分かる．

また，図 8 は「受動的戦略」,「能動的戦略」のどちらを取っても，タスクを効率的に解決することは難しい初期配置に対する被験者の振舞いを示している．

しかし，この初期配置においては，コンピュータがどんな戦略で振舞っても，被験者が取る行動戦略が定まる事がなく，データは全て同程度の確率を示している．この事から，被験者はコンピュータの振舞いに対して最適な行動が分からず，どの戦略を取ろうかと迷っていると考えられる．

4.3 人間の特性

能動的な戦略の効果がよく見られた初期配置では，被験者が Lv.1 の戦略に対して Lv.1 の戦略，Lv.2 の戦略に対しては Lv.2 の戦略というように，タスク解決が効率の良い戦略を上手く取れていない事の原因について，現段階では不明であるが，この時の被験者の振舞いを観察すると，ハンターは自己と他者が互いにどちらの獲物を狙おうかと意図の読み合いや譲り合いといった状況に陥っているケースが多かった．

しかし，これらの事例以外では，人間は我々のモデルとほぼ同等の行動をしていることが行動実験によって示された．

5 考察

我々は日常生活において，その時々の状況 (他者の戦略) によって戦略を選択している．そして，このような状況に応じた戦略選択が出来るためには，その時の状況を判断する「メタ戦略」がより上位に存在する事が計算機シミュレーションの結果から示唆された．更に，今回行なった行動実験からも，この人間における「メタ戦略」の存在を裏付ける結果を得ることができた．

計算機シミュレーションと行動実験の結果をふまえると，我々はインタラクションを実現する際，その過程には「他者の意図推定」,「自身の行動決定」,「他者モデルの適応」といった，いくつもの要素が存在する．そして，これら一つ一つの機能要素を組合せることで，我々は他者との円滑なインタラクションを実現していると考えられる．我々はこのようなインタラクション場面におけるいくつもの機能要素の組合せを「インタラクションパターン」と呼ぶ．

ここで，もし直面している状況が過去に類似した場面であるならば，その際に有効であったインタラクションパターンを選択すれば良い．また，過去に未経験の状況や選択したインタラクションパターンでは上手くいかなかった場面では，機能要素を組み合わせて新たなパターンを探索すれば良い．つまり，このような状況に応じて適応戦略を見つける過程こそ「メタ戦略」の処理と言えるのではないだろうか．

6 まとめ

本研究では，自己と他者のインタラクション過程を計算モデル化し，計算機シミュレーションによりその妥当性を評価した．我々の行動決定には受動的戦略と動的戦略が存在する．そこで，各戦略の振る舞いについて行なった行動実験の検証と，シミュレーション結果とを照らし合わせ，そこから考えられるインタラクションパターンの概念を提案した．

今後は，インタラクション場面を題材に、機能要素の組合せ探索方式の開発のため，ハンタータスク以外にも，このモデルを用いた行動実験を行ない,「メタ戦略」の存在を見出すことでこの処理過程の詳細を明らかにしていきたい．

これらの検証によって得られたデータは，今後ユーザと円滑に協調できる機械への応用に繋がっていき，更には人間の円滑な協調行動を実現する能力の理解に貢献することが出来るのではないかと考える．

参考文献

[1] Makino, T., Aihara, K.: Multi-agent reinforcement learning algorithm to handle beliefs of other agents' policies and embedded beliefs *In Proceedings of the 5th International Joint Conference on Autonomus Agents and Multiagent Systems(AAMAS'06)*, p.789-791, 2006

[2] 高野雅典, 加藤正浩, 有田隆也: 心の理論における再帰のレベルの進化に関する構成論的手法に基づく検討認知科学, 12(3), pp.221-223, 2005

[3] Nagata Y., Ishikawa S., Omori T., and Morikawa K：Computational Model of Cooperative Behavior: Adaptive Regulation of Goals and Behavior *Proceeding of the Second European Cognitive Science Conference (EuroCogSci 07)*, 202-207, 2007