

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年6月11日現在

機関番号：30106

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700293

研究課題名（和文）他者の行動価値基準推定に基づくインタラクション過程のモデル化の試み

研究課題名（英文）Behavioral Analyzing and Modeling of Interaction Processes Based on Estimation of Other's Behavior Evaluation Function.

研究代表者

石川 悟（ISHIKAWA SATORU）

北星学園大学・文学部・准教授

研究者番号：30344477

研究成果の概要（和文）：本研究では、他者とのやり取り場面において、他者の意図および行動を推定しながら自身がより柔軟に行動できるように必要な処理機能について、行動実験による分析と、処理機構を念頭においた機能モデルの提案を目指した。行動実験結果から、他者の意図や行動の推定に必要な情報の抽出と、それらの情報に基づく行動予測・意図推定とともに、推定された他者の意図に合わせ、自身の行動をあらかじめ調整し適切な行動を選択できることが明らかになった。一方機能モデルの構築には、他者から得られる情報の価値の評価機構をより詳細に検討する必要性が示された。

研究成果の概要（英文）：This study was conducted to analyze cognitive functions of interaction processes in which suitable own behavior was decided with estimated other's intention of behavior, and to construct a model of those functions. Behavioral experiments showed that participants paid attention to usable other's behaviors for estimating other's intention of behavior, and chose adequate behavior in order to fit other's intention into own intention. For functional modeling, however, it needed to assess more precise processes of interactions quantitatively, and to identify precisely evaluation processes for information value of other's behavior.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,500,000	450,000	1,950,000
2010年度	1,000,000	300,000	1,300,000
2011年度	700,000	210,000	910,000
年度			
年度			
総計	3,200,000	960,000	4,160,000

研究分野：総合領域

科研費の分科・細目：情報学，認知科学

キーワード：他者意図の推定、インタラクション、行動決定、認知科学

## 1. 研究開始当初の背景

他者とのインタラクションにおける円滑なやりとりには、インタラクション相手の行動に対して単純に応答するだけではなく、インタラクション相手の行動の意図や目的な

ど、相手の内的な状態について推定・想像することが必要である。この相手の内的状態の推定は「心の理論 “Theory of mind” (Premack & Woodruff, 1978)」あるいは「読心 “mentalizing” (Frith & Frith, 2003)」

と呼ばれ、その行動的特徴や発達機序、あるいは脳科学的・脳機能的な基盤等について、数多く研究がおこなわれているテーマである。

研究申請者はこれまで、これらのテーマの中でも円滑なやりとりを導くインタラクション相手の内的状態の推定過程について、その機能を実現する情報処理過程の計算モデルの構築に関する研究に携わった (Omori, et al., 2008; 高橋ら, 2007; Nagata, et al., 2007)。これらの研究では、インタラクションが必要な共同問題解決場面において、自身とインタラクション相手が問題場面を同一の状態として捉えていれば、自身の行動決定過程に基づいて相手の内部状態を推定し行動を予測する計算モデルが構築でき、そのモデルに従うことによって効率的に問題が解決されることが示された。一方で、この処理過程には以下の3点の前提が必要だった。

a. インタラクションの対象となる事案 (ex. 問題場面) について相手も自身も同一の状態を想定する (状態空間の同一性)。

b. 相手の実行・選択する可能性のある行動が自明 (状態行動価値関数の非推定)。

c. 相手は常に同じ価値基準でインタラクションをおこなう (行動決定態度の一定性)

実際の人間のインタラクションを考えると、インタラクション対象となる事案に対し、相手と自身が同一の状態を想定しているか自明ではなく、仮に同一の状態を想定していても実行・選択される可能性のある行動も自明でない。さらに、相手がどのような価値基準にしたがうかによって相手の行動は大きく変動するため、以前の研究で提案した計算モデルが要する前提が常に満たされるとは限らない。したがって、円滑なやりとりの実行・処理過程を実際のインタラクションに即して理解するには、人間がこれらの点を如何なる方策により解決しているのか明らかにする必要があった。

## 2. 研究の目的

本研究では、2者間のインタラクション過程についてより説明力のある機能モデルの構築に向け、その処理過程の仮説を提案すると同時に仮説の妥当性と正当性、実装可能性について行動実験を通じた検討・検証をおこなうことを研究目的とした。その際、インタラクション過程について、以下のような仮説を立てた。

行動価値基準推定モデル: 「読心」の機能には、内部状態推定と行動の予測に加え、以下の3点の推定機能が関与する。

I. インタラクション状態を定義する状態空間の推定と選択: この機能により、自身と相手との「状態空間の同一性」の前提が不要に

なる一方、インタラクション対象について状態空間を推定する過程が必要となる。

II. 推定された状態空間における行動決定のための行動価値関数の推定と選択: この機能により、「状態行動価値関数の非推定」の前提が不要になる一方、行動決定過程の推定が必要となる。

III. 行動価値関数の決定に関与する内的状態 (価値基準) の推定。この機能により、「行動決定態度の一定性」の前提が不要になる一方、行動決定のための価値基準の推定が必要となる。本研究ではこの過程を、相手の状態空間に対する態度・価値基準について推定しその推定された価値基準にしたがった状態行動価値関数の選択パラメータの決定過程、と仮定した。これにより、これまでのモデルでは扱うことが難しかった相手の行動規則が推定できずに生じるやりとりの困難さの解決や、状態が絶えず変化する相手との長期間にわたるインタラクションダイナミクスが説明できると考えた。

## 3. 研究の方法

本研究の提案仮説である3つの推定機能は、2者間のインタラクションの円滑なやりとりをおこなう複数の行動実験によって、妥当性・正当性を評価・検討した。妥当性・正当性の評価には多様な行動課題により多角的に検討する必要があるが、本研究では以下の3点に焦点を絞り行動実験課題を実施した。

(1) インタラクション当事者同士が同一の状態空間中にあっても、双方には異なる状態行動価値関数が用意され、異なった行動が生起し、インタラクション当事者は相互に行動価値関数を推定する (行動課題 A, B)。

(2) 行動を決定する行動価値関数は価値基準 (内的状態) により変化し、インタラクション当事者は相互に相手の内的状態を推定する (行動課題 A, B)。

(3) 行動を決定する状態行動価値関数は推定された内的状態により選択され、内的状態と選択された行動価値関数間に何らかの関係が存在する (行動課題 C)。

行動課題 A のインタラクション場面には、行動決定にあたり考慮すべき状態 (状態空間) が明示的に規定でき、同時に行動決定のための行動価値関数を複数設けることができると同時に、その選択には、推定された内的状態が関与すると考えられた「騙し場面」を用意した。この場面で「騙す側」が、騙される側の観察可能な状態から内的状態および行動価値関数についてどのような推定をおこなうか、そして「騙される側」によって推定されることを織り込みながら「騙す側」がどのような行動を生起させるか、検討した。

行動課題 B のインタラクション場面には、

として「共同問題解決場面」ならびに「購買場面」を用意した。両場面は、「問題解決」あるいは「物品購入」という明確なゴール状態が定義でき、ゴール状態までの状態空間や取るべき行動が比較的自明である。一方、問題の進み具合の評価や購入希望物品に対する評価など、インタラクション相手が行動決定に用いる価値基準となる内的状態は開始当初は自明ではなく、相手の内的状態の推定が必要となる。しかしながら自身が推定した相手の内的状態の適切さは直接確かめられず、相手に対する働きかけとそれに対する相手の反応から判断しなければならない。このような相手の内的状態の推定・判断が実際のインタラクション場面でもどのようにおこなわれているのか、本課題により検討した。

行動課題Cのインタラクション場面には、「説明場面」を用意した。この場面では、インタラクション当事者同士の内的状態は説明される事象の理解の進行とともに変化し、それに伴って行動価値関数も変化すると考えられる。このような継続するインタラクション場面において、推定された相手の状態と実際の相手の状態との両方が一致するよう、どのように自身の行動を変化させているのか検討した。併せて、この時の内的状態の推定過程およびその変化に伴う行動価値関数の推定と行動決定過程を検討した。

以上の行動実験結果に基づき、当初の仮説の妥当性・正当性を評価すると同時に、これらの実験結果を説明しうる機能モデルの提案をおこなった。

#### 4. 研究成果

行動課題A「騙し場面」により、以下の点が明らかになった。

まず、「騙し場面」において、「騙し」が成立することから、『(1)インタラクション当事者同士が同一の状態空間中であっても、双方には異なる状態行動価値関数が用意され、異なった行動が生起し、インタラクション当事者は相互に行動価値関数を推定する』事態であることが確かめられた。さらに、この場面において「騙す側」がおこなった行動の系列を詳細に分析した結果、「騙される側」がおこなった行動および推定内容に応じて「騙す側」は行動を随時変化させ、自身の意図を隠蔽する行動、あるいは他者の意図を誘導する行動を生起させていた。したがって、『(2)行動を決定する行動価値関数は価値基準（内的状態）により変化し、インタラクション当事者は相互に相手の内的状態を推定する』ことが部分的に確かめられた。

一方で、「騙す側」があらかじめ決められた行動系列にしたがって行動した場合、ある時点に生起させた行動が、推定されたインタラクション相手の意図に十分に合致しなかつた場合、「騙される側」の行動を変化させることができず、行動決定過程と推定される内部状態との関係が非常に強いことが示唆された。

このように、行動課題Aからは実験当初に想定した仮説と一致する結果が得られ、その定性的な評価がおこなえた一方で、処理機構を実装可能にする定量的な評価には十分至らなかった。行動分析手法の精緻化と効果的な定量的行動指標の利用によって、この点が解決できると考えられた。

行動課題B「共同問題解決場面」・「購買場面」により、以下の点が明らかになった。

両場面において、インタラクションの進行に伴って、インタラクション当事者の一方がもう一方の内的状態をどのように推定し、推定結果にしたがってどのように行動を変化させたのか、主に発話内容に着目して行動を分析した。その結果、推定された相手の内的状態にしたがって自身の発話行動あるいは身振り等の行動を調節していることが観察され、行動実験Aと同様に『(1)インタラクション当事者同士が同一の状態空間中であっても、双方には異なる状態行動価値関数が用意され、異なった行動が生起し、インタラクション当事者は相互に行動価値関数を推定する』、『(2)行動を決定する行動価値関数は価値基準（内的状態）により変化し、インタラクション当事者は相互に相手の内的状態を推定する』の2点がこの場面においても観測された。

一方で、この場面における相手の内部推定過程、および変化させた行動決定過程がどのような変数あるいは要因によって制御されていたかについては、当該場面での不確定要素および制御不可能な要素が多いことから、十分に明らかにすることができなかつた。実際のインタラクション場面により近い状況を考慮して実験を実施したが、行動課題Aのように、行動の単位および系列が明確に制御できる場面となるような工夫をおこない、そこでの処理過程が明確にできるような問題場面の作り込みが必要となった。

行動課題C「説明場面」においては、内的状態が変化する中での「説明者」の行動の決定過程を検討した。その結果行動課題A・Bと同様に、『(1)インタラクション当事者同士が同一の状態空間中であっても、双方には異なる状態行動価値関数が用意され、異なった行動が生起し、インタラクション当事者は相互に行動価値関数を推定する』、『(2)行動を決定する行動価値関数は価値基準（内的状態）により変化し、インタラクション当事者は相互に相手の内的状態を推定する』の2点が観測されると同時に、推定された内的状態の変化に伴って、柔軟に表出させる行動を変化させていることが明らかになった。さらに、「説

明者」は、推定された相手の行動の意図を自身にとってより望ましいものとなるように自身の行動を調節していた過程が明らかになった。

このことから、内的状態が動的に変化していく場面における他者意図の推定および推定に基づく自身の行動の選択、といったそれぞれの機能が行動課題Cにおいても確認されたが、この場面における内的状態と選択された行動との間の関係については定量的な評価がおこなえず、『(3)行動を決定する状態行動価値関数は推定された内的状態により選択され、内的状態と選択された行動価値関数間に何らかの関係が存在する』のうち、推定された内的状態と選択された行動価値関数との関係を定量的に記述する事はできなかった。

以上のように、本研究でおこなった3つの実験課題から、本研究当初に想定していた「読心」の機能についての仮説については、一定程度その妥当性が検討できた。しかし、この妥当性の検討は十分であるとは言えず、対立仮説（機能がない場合、あるいは機能不全となった場合）との相互検証がされていない。この機能を実装する上では、仮説として提案した機能の処理機構がどのような物であるのか、そしてこの機能が人同士のインタラクションにおいて必要不可欠な物であるのか、どのようなメリットを持つ物なのか、さらに検討が必要となっている。この点についてさらに取り組むことにより、処理過程を量的に扱うことができるようになるとともに、インタラクション過程のダイナミクスが記述できるようになることが期待される。

## 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計3件）

- ① 石川悟、自身の意図の隠蔽と他者の意図の誘導、日本認知科学会第28回大会発表論文集、査読有、2011、399-401、
- ② 石川悟、他者の行動の誘導～「騙し場面」を用いた検討～、日本認知科学会第27回大会発表論文集、査読有、2010、325-327、
- ③ 石川悟、長田悠吾、大森隆司、他者の意図を誘導する自己の行動調節～販売場面における予備検討～、日本認知科学会第26回大会発表論文集、査読有、2009、164-165、

〔学会発表〕（計6件）

- ① 石川悟、自身の意図の隠蔽と他者の意図の誘導、日本認知科学会第28回大会、2011年9月24日、東京大学、

- ② 石川悟、騙すことと見抜くこと～自身と他者への意識～、日本心理学会第75回大会、2011年9月16日、日本大学、
- ③ 石川悟、欺瞞に対する同意傾向と自己愛傾向及び攻撃性、日本心理学会第74回大会、2010年9月21日、大阪大学、
- ④ 石川悟、他者の行動の誘導～「騙し場面」を用いた検討～、日本認知科学会第27回大会、2010年9月17日、神戸大学、
- ⑤ 石川悟、長田悠吾、大森隆司、他者の意図を誘導する自己の行動調節～販売場面における予備検討～、日本認知科学会第26回大会、2009年9月10日、慶應義塾大学、
- ⑥ 石川悟、異なる意見は言いにくい～知識の違いが発言量に与える影響～、日本心理学会第73回大会、2009年8月27日、立命館大学、

## 6. 研究組織

### (1) 研究代表者

石川 悟 (ISHIKAWA SATORU)

北星学園大学・文学部・准教授

研究者番号：30344477