

他者へのかまえによる行動調整とその脳過程

Neural processing of social attitude formation

高橋 英之
Hideyuki Takahashi

大森隆司
Takashi Omori

玉川大学脳科学研究所
Brain science institute, Tamagawa University

In an interpersonal communication, we estimate an internal trait of another person from his/her superficial appearance and formulate a social attitude toward the person. In this study, we try to clarify a neural processing of this social attitude formation by behavioral and fMRI experiments using a simple competitive game “matching pennies”. These results suggest that a top-down attitude formation system exists in our brain and the function of this system is weak in autistic people.

1. はじめに

近年、家庭や公共の場で活躍するコミュニケーションロボットに注目が集まっている。特に、ロボット技術の進歩によって、見た目や動きなどは実際の人間と見間違ふようなロボットも開発されている。しかしこれらのロボットと長時間インタラクションをすると、表面上は滑らかなロボットの振る舞いに不自然さを感じてしまったり、その振る舞いがひどく機械的で単調な印象を受けてしまったりと、現状のロボットがコミュニケーションの対象としては実際の人間にはまだ遠く及んでいないことが露わになる。これは、ロボットのハード面の飛躍的な進歩に対して、そのハードを動かす知能に相当するソフト面の進歩が追い付いていないことを示している。この問題の解決への一つの道は、高いレベルのコミュニケーションを実現している人の行動決定過程を観察し、モデル化することである。そこで本研究では、競合ゲームを題材として、人間が如何に他者と円滑にコミュニケーションを行っているのかを神経科学の観点から検討することで、コミュニケーションロボットの設計に必要な知能の原理を探ることを目指す。そして本稿の最後に、その知見をうけて人工知能研究が行うべきことについて議論する。

2. インタラクションにおけるトップダウンのかまえ形成の重要性

人間のインタラクションにおけるもっとも重要な特徴の一つに、その動的な性質があると思われる。人間には直接は観測できない内部状態(心)が存在し、それが常に変化している。従って人とのインタラクションにおいてはこのような内部状態の変化を敏感に感じ取り、それに応じて自らの行動を調整しないとイケないことは明らかである。特に他者との競合場面においては、他者はこちらの行動に応じて気づかれないように自らの内部状態を変えて裏をかこうとする。その為、それに対応した“かまえ”で他者に臨むことが重要となる。逆にコンピュータなどの物理的な対象とのインタラクションでは、このような複雑な内部状態が存在することはまれである。そのため、明確なルールにもとづいた振る舞いを推定することがより有効となる。従って自分がインタラクションする相手が人間か物理的な対象化に応じて、どのような相手とインタラクションをするのか、あるいはどのような行動決定戦略を取るのかという“かまえ”をトップダウンに形成することは、円滑に対象とインタラクションする上で重要である。

他者とのインタラクションに深刻な問題を抱える自閉症の人たちは、他者に対しても物理的な対象のように振舞うことが知られている[American Psychiatric Association 2000]。これは、インタラク

ションを行う対象に応じて適切なかまえを形成するシステムに、自閉症では何らかの問題がある可能性を示唆している。

本研究では、このようなインタラクションにおけるトップダウンのかまえ形成の脳における実在を示すため、行動実験と fMRI 計測、さらに自閉症におけるその問題を検討するために高機能自閉症成人を対象とした行動実験を行ったので報告する。そしてこれらの実験の結果から、背後にあるかまえ形成の脳過程についての考察を行なう。

3. インタラクションにおけるトップダウンのかまえ形成とその自閉症における問題の実験的検証

近年、他者の存在を認識するだけで、我々の知覚や認識に様々な影響がトップダウンに変わることが報告されている[Teufel 2010] [Waytz 2010]。このような他者の存在が外界に対する我々のかまえにどのような変化をもたらすのか検討するための一つの手段として、対人ゲームを用いることは有効である。ゲームの対戦相手を他者(人間)であるかコンピュータ(物理的な対象)であるかという教示を操作することによって、実際には同じコンピュータとゲームを行っていても、教示に応じた特異的な脳活動の変化が生じることが報告されている。例えば Gallagher らは、じゃんけんゲームにおいてゲームの対戦相手を他者だと思い込むことで、相手をコンピュータだと思い込む条件と比較して前頭葉内側部の活動が増加することを PET による実験で報告している[Gallagher 2002]。また Rilling らは囚人のジレンマゲームと最後通牒ゲームという二種類の対人ゲームにおいて、ゲームの相手をヒトだと思っただけで前頭内側部に加えて上側頭溝などの活動が共通して高まることを fMRI による実験で報告している[Rilling 2004]。これらの脳部位は写真や漫画の登場人物の心の状態を推定するような課題や、幾何学図形の動きに心的な意味を付与する課題において活動が高まることが報告されており、他者の心の理解にかかわる脳部位だと考えられている[Castelli 2000] [Castelli 2002]。このように、対人ゲームにおいて対戦相手に対する教示に応じた脳活動の変化に注目した研究は数多くあるが、教示によって被験者のゲームの相手に対するかまえがどのように変化をするのか、具体的な脳過程に踏み込んだ研究はほとんどおこなわれてこなかった。我々はこのような対戦相手に対する教示に応じた脳活動の変化が、ゲームの相手に対するかまえの形成を反映しているのではないかと考え、ゲームの相手が他者だと教示することによる行動の変化に注目した実験を行なった[Takahashi 2007]。具体的には、硬貨合わせ課題という単純な対戦ゲームをコンピュータプログラム上で行う課題を用いた。この課題は二人のプレイヤーが、二つの行動の選択

肢から一つの行動を選択し、双方の選択の組み合わせに応じてそれぞれの勝敗が決まるゼロサムゲームである。被験者にはこのゲームを毎試行、金銭報酬を賭けて繰り返し行なってもらい、獲得する報酬を可能な限り増やしてもらうよう教示した。そしてコンピュータプログラム上で被験者の勝率を統制した上で、他者かコンピュータかというゲームの相手に対する認識を教示により操作することで、被験者の行動や脳活動がどのように変化するか調べた。

行動指標として、被験者の行動がどれだけ過去の被験者と対戦相手の行動から独立しているか(すなわち非定型的か)を被験者の行動選択の条件付き確率のエントロピーから計算した。エントロピーの値は行動選択のランダムさを反映し、その値が大きいほど、被験者の行動選択は非定型的であり、エントロピーの値が小さいほど型にはまった行動選択を被験者が行っていると言える。このゲームでは、被験者と対戦相手の選択した行動に依存して決定論的に被験者の勝ち負けが決まる。すなわち被験者の行動選択の非定型的性を評価するエントロピーは、同時に被験者がどれだけ対戦相手の行動選択が非定型で予測が困難な対象であると認識していたのかを示す主観的予測性の指標となる。

実験では、ゲーム勝率を 50%に統制した上で、被験者に対する対戦相手の教示をヒトとコンピュータと交互に変化させ、教示に応じて被験者の行動選択のエントロピーがどのように変化するかを検討した。具体的には 50 試行を 1 ブロックとして、ゲームの相手がヒトであると教示するブロックと、コンピュータであると教示するブロックを交互に 3 回ずつ、計 6 ブロック行った。知能指数が高く言語能力に問題がない自閉症スペクトラム障害成人 19 名と年齢をそろえた健常成人 19 名に被験者として参加してもらった。実験結果としてそれぞれの被験者群ごとの各ブロックにおける被験者間でのエントロピーの平均は図 1 のようになった。この図が示すように、実際にはゲームの相手は常に勝率 50%のコンピュータであったにもかかわらず、健常成人ではゲームの相手に対する教示に応じてエントロピーの値が変化、すなわちゲームの相手がヒトであると教示した場合の方が、コンピュータであると教示した場合と比較して有意にエントロピーが大きくなった。しかしこのような教示に応じたエントロピーの変化は、自閉症スペクトラム障害成人ではみられなかった。

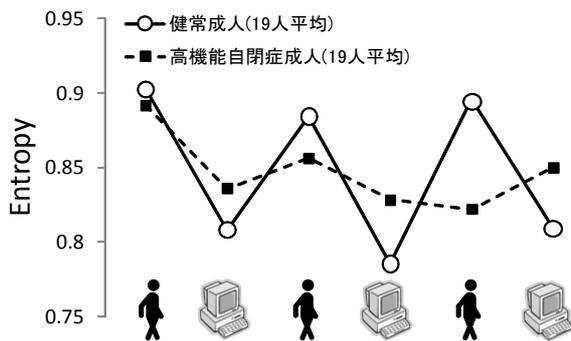


図 1. 健常成人被験者と自閉症スペクトラム障害成人被験者の硬貨合わせ課題におけるゲームの相手の教示に応じたエントロピーの変動

我々はこのようなゲームの相手に対する教示に応じたエントロピーの変化を説明する計算モデルとして、ゲームの対戦相手の振る舞いの非定型性に対するトップダウンの見積もりをパラメータとして組み込んだ強化学習モデルを提案した[高橋 2008]. こ

の計算モデルにもとづく硬貨合わせ課題のシミュレーションにより、ゲームの相手に対する予測性(見積もり)のパラメータをゲームの相手の教示に応じたかまえとして調整することで被験者のエントロピーの変動が説明できること、具体的にはゲームの相手を他者と教示した条件の方がコンピュータと教示した条件よりも予測性(パラメータ)の見積もり下げることによって行動実験における健常被験者と同じエントロピーの傾向が得られること、さらにこのようなパラメータのかまえとしての調整が行われていないと仮定することで自閉症スペクトラム障害被験者の行動データをうまく説明できることを示した。

このようなかまへの形成メカニズムの背後にある脳内基盤を探るために、fMRI を用いて同様の課題を行っている健常被験者の脳活動を計測した[Takahashi 2010]. その結果、まずゲームの相手が他者であると教示された場合、コンピュータであると教示された場合でそれぞれ異なる脳部位に活動がみられた(図 2). 具体的には、前者では前頭葉内側部や上側頭溝などの先行研究と同様の脳部位、また後者では紡錘状回や側頭平面などの脳部位に活動がみられた。これらの脳活動はゲームの対戦相手の教示に応じた特有の脳活動であると言える。

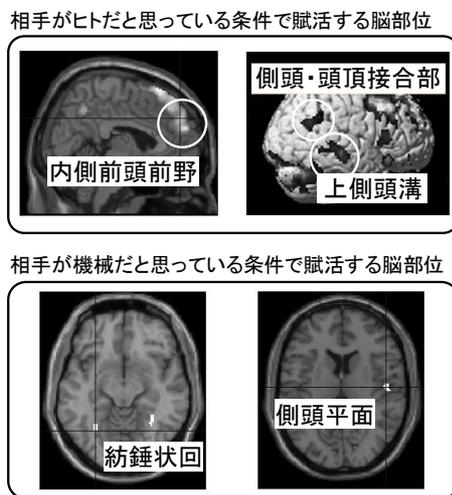


図 2. ゲームの相手の教示に応じた脳の賦活部位

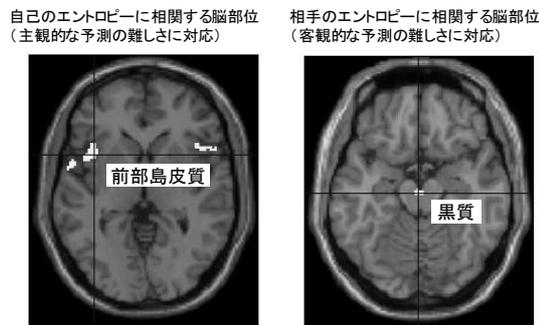


図 3. 自己のエントロピー(主観的な予測の難しさ)に対応する脳部位(左)と他者のエントロピー(客観的な予測の難しさ)に対応する脳部位(右)

それに対して、ゲームの相手についての教示とは別に、ゲームの相手に対するかまえを反映する脳部位を調べるため、被験者の行動選択のエントロピーと相関する脳部位と対戦相手の行動選択のエントロピーと相関する脳部位を調べた。前者は被験

者がゲームの対戦相手に対して感じる主観的な予測の難しさを反映する脳部位と考えられ、後者は客観的な対戦相手の予測の難しさを反映する脳部位であると考えられている。主観的な予測の難しさに対応する脳部位として、背外側前頭前野や頭頂間溝などに加えて前部島皮質に活動がみられ、客観的な予測の難しさに対応する脳部位として脳幹の中脳に活動がみられた(図3)。

4. インタラクションにおけるトップダウンのかまえ形成のモデル化

これまでの我々の検討から、ゲームの対戦相手に対する思い込みがトップダウンに被験者の行動を調整することが示された。これは「ゲームの対戦相手についての認識がより下位の行動選択のシステムの構えを形成する」というタスクセットの枠組みで捉えることができよう。

硬貨合わせ課題において人が行動選択に至るプロセスは、過去の被験者および相手の行動履歴が状態として認識され、認識した状態にもとづき相手の行動予測を行い、予測に応じた行動判断を行うという点については対戦相手がどのようなものであれ共通であると思われる。対戦相手への思い込みがトップダウンに影響を与えるとすれば、それは被験者が有している対戦相手の行動モデルの影響であろう。このゲームのプレイヤーは一般に、対戦相手の行動を予測するために相手についての何らかの行動モデルを脳内に形成し、それにもとづき相手の行動を予測すると考えられる。どのようなモデルを対戦相手に仮定するかは、被験者がヒトやコンピュータに対して持つメンタルモデルに依存する。一般には、相手がヒトだと思込むことで、対戦相手の中にも被験者の行動モデルが形成されているという、入れ子構造のモデルが仮定され、逆に相手がコンピュータだと思

い込むことで、相手が何らかのルールに従って行動選択をするという因果的なモデルが仮定されると考えられる(図4)。

このような相手に応じた脳過程の違いは、相手の行動の直接的な予測以外にも、より大局的な被験者の行動に影響を与えると考えられる。すなわち、相手がこちらの行動を予測して行動を変えてくると仮定した場合、被験者の対戦相手に対する不確実性(リスク)の知覚が大きくなり、その結果として相手に行動を読まれることを回避するために自らの行動規則を探索的、すなわち非定型的(エントロピー大)にすると考えられる。逆に、相手が何らかの行動規則に従って行動を選択していると仮定した場合、被験者の対戦相手に対する不確実性(リスク)の知覚が小さくなり、その結果として相手の行動規則に従った利用的、すなわち定型的(エントロピー小)な行動をとると思われる。実際に、実験に参加した多くの被験者がこのようなスタンスで行動選択を行なったと内省報告している。

5. まとめ

以上、相手に応じたかまえ形成の脳過程について、我々が行ってきた実験結果にもとづいて現在想定しているモデルについて述べた。脳がこのような相手に対するかまえを取ようになったのは、それが有効に働くからであり、当然のことながら我々の日々の経験の蓄積がその理由となっているはずである。すなわち、相手についての認識にあわせて自己の行動決定方略をかまえとして変化させることは有効であるし、統計的に意味のある根拠があるものと考えられる。

この知見を人工知能におけるコミュニケーションに置き換えるなら、以下のことが言えよう。

1. コミュニケーションには他者の心的状態の推定を含む行動決定モデルが必要である。他者の心的状態は警告なしに変動するため、コミュニケーションにはたえずその変動に伴う不確実性がつきまとう。それに対応した行動決定モデルの開発が必要である。それが無い場合、自閉症者のコミュニケーション障害のような問題が発生するであろう。
2. 行動決定モデルは、他者の特性に応じて切り替える必要がある。我々は日々のコミュニケーション活動において、他者の特性に応じて自己の行動決定過程を調整しているの考えるのが自然である。

しかし一方で、他者の心的状態の推定に関しては多くのことは知らない。横山らや長田らは自己の行動決定モデルを他者に投射することで他者の意図推定をする方式を定式化し、さらに他者が自己の意図推定をする場合も含めた行動決定モデルを検討している[横山 2009][長田 2010]。しかし現時点での人間のコミュニケーション場面での行動決定については不明確なことが多く、その計算モデル化は今後の課題である[横山 2011]。

参考文献

[American Psychiatric Association 2000] American Psychiatric Association. Diagnostic and statistical manual of mental disorders (4th edition, text revision) DSM-IV-TR, 2000.
 [Teufel 2010] Teufel, C., Fletcher, P. C., & Davis, G. Seeing other minds: attributed mental states influence perception. Trends in Cognitive Sciences, 14, 376-382, 2010.
 [Waytz 2010] Waytz, A., Gray, K., Epley, N., & Wegner, D. M. Causes and consequences of mind perception. Trends in Cognitive Sciences, 14, 383-388, 2010.

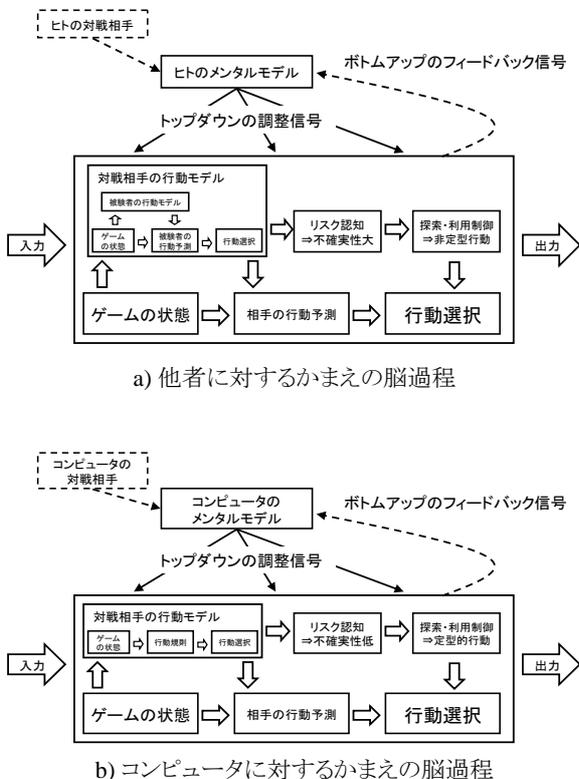


図4. 相手に応じたかまえの脳過程

- [Gallagher 2002] Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. Imaging the Intentional Stance in a Competitive Game. *NeuroImage*, 16, 814-821, 2002.
- [Rilling 2004] Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, 22, 1694-1703, 2004.
- [Castelli 2000] Castelli, F., Happé, F., Frith, U. & Frith, C. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12, 314-325, 2000.
- [Castelli 2002] Castelli, F., Frith, C., Happé, F. & Frith, U. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, 125, 1839-1849, 2002.
- [Takahashi 2007] Takahashi, H., Samejima, K., Omori, T., Murohashi, H., & Kamio, Y. Can individuals with autism pre-modulate their decision making process depending on social context? *Neuroscience 2007, SfN's 37th annual meeting, USA, 2007.*
- [高橋 2008] 高橋英之・石川悟・大森隆司:円滑な対人インタラクションを実現する対象認識に応じた認知的構え調整機構のモデル化, *認知科学*, 15, 202-215, 2008.
- [Takahashi 2010] Takahashi, H., Izuma, K., Matumoto, M., Matumoto, K., & Omori, T.. Neural correlates for strategic adjustment during interpersonal competitive game. *Proceeding of 16th Annual Meeting of the Organization for Human Brain Mapping. Spain, 2010.*
- [横山 2009]横山絢美, 大森隆司:協調課題における意図推定に基づく行動決定過程のモデル的解析, *電子情報通信学会論文誌 A, Vol.J92-A, No.11, pp.734-742, Nov.2009*
- [長田 2010]長田悠吾, 石川 悟, 大森隆司, 森川幸治:意図推定に基づく行動決定戦略の動的選択による協調行動の計算モデル化, *認知科学, Vol.17,No.2,pp.270-286,2010*
- [横山 2011]横山絢美, 大森隆司:対人コミュニケーション場面における人の行動決定過程のモデル化とその評価, *A-9, 脳と心のメカニズム第 11 回冬のワークショップ, 2011*