

題目：ヒトの心を汲みとる社会知性——脳神経基盤と脳計算の解明——

Title: Reading others' minds: pursuit of social brain functions from computational perspective

中原裕之

Hiroyuki Nakahara

理化学研究所 脳科学総合研究センター 理論統合脳科学研究チーム [〒351-0198 埼玉県和光市広沢 2-1]

Laboratory for Integrated Theoretical Neuroscience, Riken Brain Research Institute, 2-1 Hirosawa, Wako City, Saitama 351-0198, Japan

はじめに

私たちの日常生活の大半は何らかの社会行動である。つまり、「ヒトである」ことの多くは、その社会行動のなかに表れるといえる。したがって、この社会行動をつかさどる神経基盤の解明は、すなわち「ヒト」を理解することにつながるといえるだろう。脳神経基盤の解明は脳科学の重要目標であると同時に、広汎な諸科学の発展をうながす壮大な目標でもある。脳科学の発展は著しい。百年前、あるいは50年前、10年前に比べると、実験技術の進展は目覚ましく、新たな知見も急速に増している。とはいえ、この究極の目標までの道のりは、今はまだ遠いかもしれない。しかし、筆者は、いつかそこにたどり着くことができると信じている。

ここで、筆者の構想をまず述べておきたい。社会行動の根幹には、人々が互いの心や行動を推断する社会知性がある。この社会知性の脳メカニズムを、大胆な捨象を経て「自己システム+他者システム」の観点から解明する。それには実験と理論の融合研究が必要で、社会知性の脳計算の基本構成要素を徹底的に明らかにするのが最初の重要な一歩となる。このような話をされても何のことやらさっぱり、と思う方もいるだろう。本稿では、この構想の源流を述べたうえで、筆者らの最近の研究成果を紹介し、今後の展望について簡単に述べたいと思う。

現代は「脳内他者」を研究できる時代

私たちの日常の社会行動や、その行動をとる際に私たちの心に去来するさまざまな想いや感情、思考について考えると、その複雑さと奥深さにはめまいさえ覚える。そもそも私たち研究者は、その複雑さにこそ惹かれているのかもしれないし、その複雑さを追求する

のも、もちろん一つの道ではある。しかし、筆者は、まずは複雑さを切り捨てて、その根幹を探究することに最初は焦点をあてるべきと考えている。社会行動の根幹には、他者の心や行動を推断できる社会知性がある。この社会知性の本質は、言うなれば“脳の中に他者がいる”ことだ。たとえば「心の理論」は、脳科学のみならずヒト社会知性に関わる諸学問の中心的な論点である(たとえば、[1-3])。それは、端的には「他者の心(状態や意図など)の推測と、それにより他者の行動を予測し、それを自らの行動に生かす」という脳機能のことである。つまり、この「脳内他者」を明らかにしていくことが重要だと筆者は考えている。

私たちは現在、人の行動と脳活動を同時に観察できる時代に生きている。筆者の場合は、ヒト非侵襲脳活動イメージング実験(磁気共鳴機能画像法、functional magnetic resonance imaging; 以下、ヒトfMRI)を用いている。古来より文学書や歴史書、哲学書など幾多の書物に見られるように、現代にいたるまで広汎な社会科学において、社会知性に関してさまざまな議論や研究がなされてきている。もちろん、先人・偉人の知見には大いに学ぶべきであるが、彼らは私たちと同じ時代に生きていたわけではない。たしかに、最新機器とはいえ、ヒトfMRIにもその時間・空間解像度にはまだまだ限界がある。したがって、これがあればすべて解決するというわけでもない。しかし、あえて繰り返すが、それでも今は、先人の生きた時代とは違う。彼らの社会知性に対する問いと答えを、現代の私たちは「脳内他者」という新たな問いかけとして検証することができるのである。

脳計算の基本構成要素

脳の機能を理解するとは一体どういうことだろうか?そもそも何が分かればよいのだろうか?ここで忘れてならないのは、脳では物質原理と情報原理が交差するということだ。脳における多数の神経細胞集団の活動が脳機能を生み出している。多様な種類の神経細胞はもとより、種々の神経伝達物質やその受容体、あるいは複雑な脳回路構造、そして、分子・遺伝子レベルの多様性もある。これらの物質原理が脳には存在する。一方で、これらの物質原理に支えられた神経細胞集団活動のパターンと相互作用こそが、私たちが日常用語で言うところの、たとえば知覚・認知・運動であり、他者の心なのである。それは物質性を超えて(あるいは含めたうえで)、脳の情報処理、すなわち脳計算の産物である。脳計算という捉え方は、脳の物質原理と情報原理の交錯を明確にする視点である。脳計算という視点なしでは「脳内他者」として社会知性を理解することはできないだろう。

さらに、脳計算という視点は、行動と脳を結びつけるためにも必要不可欠である。そもそも脳機能の理解、つまり脳情報処理の理解は、行動と対応づけて行われる必要がある。これは脳に限った話ではない。たとえば、何かのコンピュータゲームを想像して、そこに出てくるキャラクターの複雑な行動を思い浮かべてみるとよい。その複雑な行動とそれを生成する機構(この場合は、コンピュータ上のプログラム)を理解するには、その生成機構

の作動目的を理解し、そのアルゴリズム自体を理解する必要がある。D.Marr[4, 5]が喝破したように、脳機能を理解するために必要な3つの水準、①計算理論（作動目的）、②表現とアルゴリズム、③ハードウェアによる実現、のそれぞれを的確に尊重しなくてはならない。対象とする行動が複雑であればあるほど、行動と脳を結びつける脳計算の観点はより一層重要になる。その行動の生成機構を理解するのに、その作動目的とアルゴリズムの理解がより重要になるからである。社会知性の根幹の「脳内他者」を理解するには、脳計算を一層重視して研究を進めなくてはならない。

では、脳計算を明らかにするにはどうすればよいのか？千里の道も一歩から、というように、ここで大切なのは、基本構成要素という視点である。これは、多様な、あるいは複雑に構成された計算（情報処理）の土台になる要素である。たとえば、現代は、空港や商店街などさまざまな場所に無人カメラが設置され、画像が自動的に撮影される時代である。それらの画像に自動認識アルゴリズムを用いると、人物の認識などが可能である。そのような状況は、サイエンス・フィクションを除けば、1950年から1960年に入るところには想定しづらかったのではないか。じつはそのころ、後にノーベル生理学・医学賞を受けた Hubel & Wiesel[6, 7]は、初期視覚野の方向選択性神経細胞活動を発見していた。筆者は、この発見が自動認識アルゴリズムにつながったと考えている。もちろん、彼らの発見から自動認識アルゴリズムに一足飛びに発展したわけではなく、その間には多くのさまざまな知見の積み重ねがあった。自動認識アルゴリズムにもまだ向上させるべき点はあるし、物体認識の脳機能に関して未解明の点も多い。しかし、それでも振り返ってみれば、彼らは物体認識の脳計算の基本要素を発見していたといえる。それが、その後の物体認識の脳研究の発展につながり、計算（情報処理）面では自動認識アルゴリズムにつながった。このことから、複雑さに惑わされずに脳計算の基本となる構成要素を徹底解明することが、いかに重要かを痛感させられる。

「自己システム+他者システム」

「脳内他者」とは、言い換えれば、脳の中に存在する、他者の心や行動を推測する「他者システム」である。その脳内他者は、他者の心や行動の推測を実行し、脳すなわち「私」が適切な社会行動を実践できるようにしている。脳は、もともと（非社会的状況でも）存在する「自己システム」をもっている。したがって、他者の心や行動を推測し、それを生かす社会知性の脳機能を理解するには、「自己システム+他者システム」という観点が重要になる(図1)。

この「自己システム+他者システム」の観点からの研究を加速させるには、まずは「自己システム」の脳計算の基本要素を明確にすることが大切だ。そこで筆者は報酬予測に関連した社会的意思決定に着目した。その理由を以下に述べる。

報酬予測に基づく意思決定は、ヒトひいては生物一般にとって本質的な機能である。私

たちの日常は意思決定の連続である。そして、意思決定は未来に影響を与える。逆に言えば、未来予測が意思決定の際には働く。予測の中でも、報酬の予測はヒトや動物にとって原初的な部類に属し、動機づけ行動の土台となる[8-11]。そのため、報酬予測に基づく意思決定（報酬予測は、価値と呼ばれ、通常、価値に基づく意思決定：value-based decision making という；以下、「価値意思決定」）における脳機能の解明は、脳科学の重要なテーマとなっている。

この価値意思決定については脳計算からの機能理解が非常に進んでいる。そもそも、この分野の発展は報酬予測の脳計算理論（強化学習：reinforcement learning theory）に支えられてきた(図2)。この発展で特筆すべきは、脳計算の基本要素、すなわち、各選択肢の価値、さらにそれらの価値の比較に基づく行動選択確率、また、行動を選択した結果得られる報酬と予測された報酬の差から生成される報酬予測誤差が、学習信号として明確化されたことである。これらの基本要素の明確化が、さまざまな脳領域の脳活動機能を理解する上で大きな助けとなった。さらには、より複雑な価値意思決定の脳計算の明確化にもつながり、それが脳活動や脳回路のさらなる理解を促している。筆者もささやかながらこの分野の発展に貢献してきた[11-14]。ただし、この価値意思決定の脳機能理解は、主として非社会的状況、すなわち他者のいない状況での理解である。

この価値意思決定についての構成的な(脳計算の)理解を「自己システム」に据えることで、社会知性の脳計算を理解するための第一歩を進められる。この「自己システム」の土台のもとで、「自己システム+他者システム」として社会的意思決定の脳計算の解明を進めればよい、と筆者は考えている。

脳内シミュレーションによる、他者の心の学習

以上、本研究の構想を簡単にまとめてみたが、ここで、この考え方に基づいて行った直近の研究の成果を紹介したい[15] (図3)。「心の理論」の諸説のなかには「心のシミュレーション」という有力な考え方がある。これは、他者の心を理解あるいは予測するには、他者の心的過程（つまり、他者の脳で起きている脳計算）を、同様の脳計算をする自己の脳部位を使ってシミュレーションすればよいというものである。本研究では、この「心のシミュレーション」に真っ向から挑んだ (図2)。その脳計算の基本構成要素を重点的に抽出するための実験パラダイムを開発し、脳数理モデルを構築し、それをヒトの実際の行動とそのときの脳活動に照らして定量的検証を行った。その際、モデル化解析と呼ばれる理論と実験の融合研究に適した手法を用いたが、紙面の都合からここではその説明は割愛する[16]。

この実験のコントロール課題は、標準的な価値意思決定課題（自己の報酬予測による選

択)であり、ここでは非社会的状況における「自己システム」の脳計算について調べた。一方、メイン課題は、他者の価値意思決定予測課題であった(他者がコントロール課題をしていて、被験者はその選択を当てる課題)。これにより、他者の価値意思決定をシミュレーション学習する脳計算が明らかになった。この研究では、実は「自己システム+他者システム」の限局した場合を扱っている。たとえば、このメイン課題では「他者システム」の“出力”、つまり他者の行動予測を被験者は直接答えている。そこでは、日常生活なら頻繁に起こるだろう、その行動予測をもとにした自己の行動選択の再調整などは入っていない。社会性で議論になる自己と他者の間の相互作用も注意深く排除した。このように「自己システム+他者システム」を意図して限局的に扱ったのは、脳計算の基本要素を徹底的に解明するためである(現在、これを発展させた研究をしている。機会があればいつか触れたいと思う)。その理由を視覚の例で説明すれば、まずは、“物体認識”よりも、“方向選択性”にあたる基本要素の解明が必要と考えたからである。

本研究により、2つのシミュレーション学習信号が発見された。それは「他者の報酬予測誤差」(他者が実際に得た報酬量と、シミュレーションによる他者の予測報酬量の差)と「他者行動予測誤差」(他者の実際の行動と予測された行動の差)である。この他者報酬予測誤差は、全脳の中でほぼ前頭葉腹内側部のみで表されていた。一方で、他者行動予測誤差の場合は、前頭葉背外側部や背内側部、側頭頭頂接合部、後部上側頭溝でも脳活動が見られた。端的に述べれば、実際の行動に一致する脳活動の実体(脳計算)と対応する、新たな心のシミュレーション学習のモデルを提示できたと考えている。興味を持たれた読者は、概説や原著論文を参照されたい[15, 16]。

まとめ

ヒトの心を汲みとる社会知性の脳機能の解明は、将来の学問である人間総合科学の基礎となるだろう。脳計算の観点から理解することで、個々人の社会知性の違いや、ひいてはその障害や疾患の理解、そして計算精神医学という新分野の勃興を促す。脳計算という視点は、脳の物質原理と情報原理の交錯の要になる。また、行動と脳活動をつなぐ、情報処理という具体性をもった言語を生む。その言語の普遍性は、計算機科学や統計情報科学などを含む数理諸科学や工学と、さらには、文化、社会、経済、心理などを含む人文・社会科学とを包括する人間総合科学の形成において、諸学問にまたがる共通言語を提供することになる。

社会知性の脳機能を視覚・運動・認知機能と同様に理解するには、脳計算の基本構成要素をまずは徹底的に解明しなければならない。筆者らはそれを目ざしている。この挑戦がいのある課題にとともに取り組む仲間が増えることを大いに歓迎したい。社会知性の脳科学の新しい教科書が書かれる日も、じつはそれほど遠くないかもしれない。

引用文献

1. Adolphs, R.: *Neuron* **65**(6): 752-767, 2010
2. Sanfey, A.G.: *Science* **318**(5850): 598-602, 2007
3. Frith, C.D. and U. Frith: *Science* **286**(5445): 1692-1695, 1999
4. マー, デ.: *ビジョン—視覚の計算理論と脳内表現*. 産業図書. 1987.
5. Marr, D.: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*: The MIT Press. 1982.
6. Hubel, D.H. and T.N. Wiesel: *The Journal of physiology* **160**: 106-154, 1962
7. Hubel, D.H. and T.N. Wiesel: *The Journal of physiology* **148**(3): 574–591, 1959
8. Schultz, W.: *Journal of Neurophysiology* **80**: 1-27, 1998
9. Rangel, A., C. Camerer, and P.R. Montague: *Nature Reviews Neuroscience* **9**(7): 545-556, 2008
10. Hikosaka, O., K. Nakamura, and H. Nakahara: *Journal of Neurophysiology* **95**(2): 567-584, 2006
11. Nakahara, H.: *Current Opinion in Neurobiology* **25**: 123-129, 2014
12. Nakahara, H. and S. Kaveri: *Neural Computation* **22**(12): 3062-3106, 2010
13. Nakahara, H., K. Doya, and O. Hikosaka: *Journal of Cognitive Neuroscience* **13**(5): 626-647, 2001
14. Nakahara, H., H. Itoh, R. Kawagoe, et al.: *Neuron* **41**(2): 269-280, 2004
15. Suzuki, S., N. Harasawa, K. Ueno, et al.: *Neuron* **74**: 1125-1137, 2012
16. 中原裕之 and 鈴木真介: *BRAIN and NERVE* **65**(8): 973-982, 2013

Figures

Fig 1

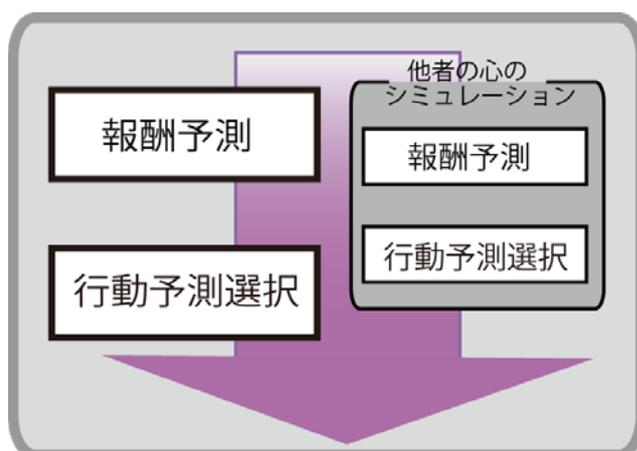


図 1 「自己システム+他者システム」

Fig 2

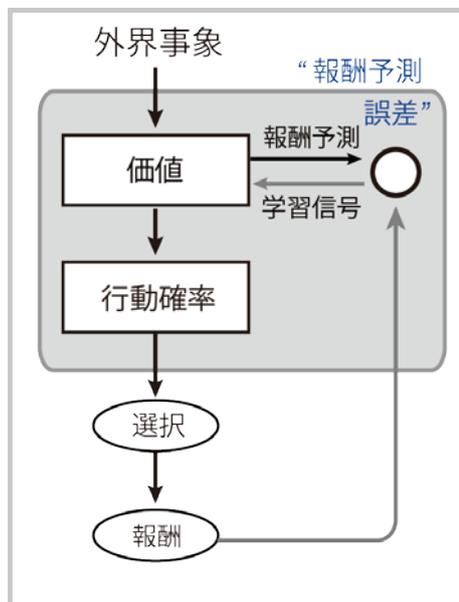


図 2 価値意思決定の脳計算の基本構成要素(強化学習)

外界入力に基づき、各選択肢に基づく報酬予測(価値)が形成される。その価値の比較から行動選択確率が形成され、それに基づき行動が選択される。その結果得られた報酬から算出した報酬予測誤差(実際の報酬量と予測報酬量の差)により、その選択肢の報酬予測が学習(更新)される。このプロセスにより、経験を積む中で、同じ状況下で適切な報酬予測および行動選択ができるようになる。

Fig 3

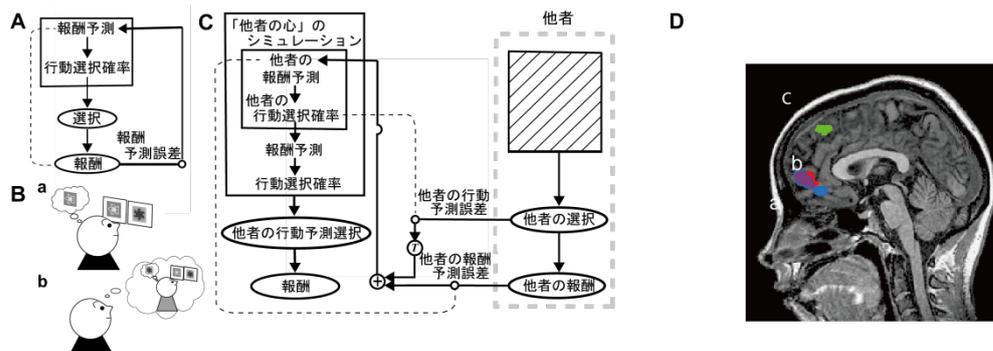


図3 他者の心のシミュレーション学習 (Suzuki et al., 2012)

A. 価値意思決定の脳内過程: 四角の枠は個体の内部プロセスを表す。B. (a) 価値意思決定の実験課題。(b) 他者の価値意思決定予測課題。C. Bの(b)の課題で他者の意思決定をシミュレーション学習する脳内過程: 2つの学習信号である「他者の報酬予測誤差」(他者が実際に得た報酬量とシミュレーションによる他者の予測報酬量の差)と「他者の行動予測誤差」(他者の実際の行動と予測された行動の差)の存在と脳機能局在を示した。D. 主な脳活動: 価値意思決定課題を遂行中の脳活動部位(a)と他者予測課題を遂行中の脳活動部位(b, c)。a: 被験者自身の「報酬予測誤差」(前頭葉腹内側部: 画像の青色部分)。b: 「他者報酬予測誤差」(前頭葉腹内側部: 画像の赤色部分)。c: 「他者行動予測誤差」(前頭葉背内側部: 画像の緑色部分)。