

# 科学

Science Journal *KAGAKU*

VOL.71 2001 別刷

---

---

岩波書店

---

ヒト知性の計算神経科学(第6回:最終回)

# ヒトの知性に計算理論はどこまで迫られるか

最終回である第6回では、連載を締めくくって、ヒト知性の計算神経科学の目指すところ、難しさ、方法論、そして基本となる計算理論を解説する。最後に、今後の課題を検討する。

川人光男 銅谷賢治 春野雅彦

かわと みつお どうや けんじ はるの まさひこ

ATR人間情報科学研究所  
ATR人間情報科学研究所, 科学技術振興事業団・戦略基礎研究・脳を創る  
ATR人間情報科学研究所

## 脳機能の理解とは何か

脳の働きがわかるというのはどういうことだろうか。脳の特定の場所がある特定の機能にとって重要な役割を果たしているという知識は大切である。例えば、大脳皮質のブローカ野(左の前頭葉にあるブロードマンの44,45野)が、言語を産出することに、なかでも文法機能に関係しているらしいということは、言語の脳科学、さらには大脳皮質の機能局在研究の中でも、150年の歴史をもつ大発見である。場所に関するこのような知識は、損傷されたヒトの脳の研究、動物の脳での局所破壊実験、さらに最近では脳活動の非侵襲計測実験によって、すばらしい勢いで蓄積されている。

また脳の特定の働きに欠くことのできない物質に関する知識も、場所に関する知識以上に急激に蓄積されつつある。細胞生物学などの分野で開発された分子生物学的な実験技術や知識体系が、脳研究に系統的に応用されて大成功をおさめた結果、分子神経生物学はこの30年で脳科学の一大分野となった。例えば、現象としては20年前に発見された小脳のプルキンエ細胞のシナプス可塑性である長期減弱の分子レベルのメカニズムが解明されつつある<sup>(1)</sup>。

脳機能の研究手法としてはニューロンの発火と行動の相関を調べるなどのシステム神経科学が最

も重要である。しかしながら、行動中のサルから単一ニューロンの電気的活動を記録して、それと刺激や運動のあるパラメータの間に相関を発見しても、それだけでは脳機能の正しい理解につながるとは限らない。その良い例は、Georgopoulosらの、運動方向のポピュレーション符号化仮説であろう<sup>(2)</sup>。この仮説では、大脳皮質運動野の多数のニューロンは、3次元空間の中での手先の特定の運動方向を表現していると考えられる。これは‘Science’の表紙などを何度も飾り、教科書にも頻りに引用される有名な研究であるが、最近の実験的、計算論的研究によって、本質的には誤っているらしいことがわかってきた。大脳皮質運動野の多数のニューロンは筋肉の活動度を符号化しているが、ある実験条件では、それと運動方向に相関があるだけなのである<sup>(3)(4)(5)</sup>。運動制御に伴う逆ダイナミクスや逆キネマティクスの計算問題としての難しさを過小評価し、神経回路の同定を行わず、概念的な仮説に基づく運動のある側面とニューロン発火の相関だけに依存した研究手法が危ういことを示している。

脳について場所、物、ニューロン発火と現象との相関だけを調べるアプローチは、一つ間違えば、宇宙人が地球上のテレビを調べた次のようなとんちんかんな研究のようになってしまう危険性がある。たとえば、テレビのプラグを抜いたら映らな



なくなったとか、ブラウン管を金属バットで壊したら煙が出てその後映らなくなったとか、テレビの電源部と家の外の送電線の周波数には相関があるなどという知識は、皆テレビに関して事実ではあるが、テレビの機能や機構の理解からはほど遠い。

場所、物質、相関に関する脳の知識は驚くほどの勢いで拡大しているが、それだけでは、脳の機能の理解にはただちにつながらないことも、これらの知識が蓄積すればするほど、かえって明らかになってきた。例えば、視覚、運動制御、言語など、私達人間が日常生活で難しいとも思わずに解いている問題は、計算機やロボットにとって相変わらずとてつもなく難しく、ほとんど実現できていない。これらの機能に関する脳の働きが本当にわかっているならば、それを人工的な機械に移植することは原理的に難しくはないはずである。それが実現していないということは、私達が相変わらず、脳の機能、さらに言えば情報処理に関してほとんどなにも知り得ていないことを示している。このような状況を打破するために、脳の計算論的な研究が必要であることが認識されたのは比較的最近のことであるが、ここ10年くらいで、いくつかのめざましい成果が得られるようになってきた。

計算論的神経科学は、次のように定義できる。脳が解いているのと同じ問題を、脳が用いているのと同じ原理で解くことのできる、計算機のプログラム、もしくは人工的な機械を作れる程度に、脳の機能を深く理解するアプローチである。これによって、旧来の脳科学が陥る危険性のある、単なる場所、物質、表面的な相関に関する知識の積み重ねの学問からの脱皮を計っているのである。キャッチフレーズとしては、脳を創ることによって、脳を知るのである。また、脳を創れる程度に脳を知ることを目指すのである。

### 計算論的神経科学の方法論と必要性

理論の予測を実験的に客観的に検証できるという意味での厳密科学としての計算論的神経科学の研究は難しい。自然科学ではないと割り切り、脳との厳密な対応をとることはあきらめて、数学的な美しさだけを求めて神経回路モデルを研究する

のとは立場が違う。

計算理論の予測を、実験で客観的に検証できるためにはいくつもの必要条件が満たされていなければならない。まずある脳機能について、脳だけではなく、身体と環境の物理学を理解して、機能を実現するための計算論的困難を明らかにし、それを解決する理論を発見しなくてはいけない。一方、実験的には、その脳機能を司る神経回路が、包括的にまた緻密に解析されている必要がある。関連する脳部位の解剖学的結合が明らかにされ、各部位の破壊実験が行われ、電気刺激と電気記録等による神経生理学的手法で回路が同定され、さらにトレーサーなどによってニューロン間の局所回路まで明らかにされている必要がある。また、脳が特定の機能を果たしているときの、脳の各部位のニューロンの発火頻度の刺激選択性、潜時、波形等が解析されている必要がある。このような条件が整って初めて、計算理論を実験的に検証することが可能になる。また計算理論の側も、単にこれらのデータがそろうのを待っているのではなく、理論から見て重要な実験を企画し提案することが求められる。しかし、そのためには、実験的に何が可能で何が困難かを理解して、理論が実験的に効率よく検証できるように、理論とそこから導かれる予測をとぎすまさなくてはならない。

追従眼球運動と平滑性追跡眼球運動に関わる小脳や大脳皮質 MST 野の実験的研究と<sup>(6)(7)</sup>、これに基づく小脳と大脳の内部モデル仮説の検証は、これらの条件が満足されていた<sup>(8)(9)</sup>。本連載にとって重要な計算論の成果には、上にも述べた大脳皮質運動野の研究、視覚野の研究、無脊椎動物の中枢パターン生成機構、大脳基底核と強化学習などがある。

解剖学的、電気生理学的ないわゆるシステムレベルの脳研究は、技術的に非常に難しく、手間と時間がかかり、しかも理論的な研究も必要な、ある意味で論文生産効率の悪い研究である。なぜ計算論的神経科学とシステム神経科学のような、労多くしてその割には報われることの少ない困難な研究が必要なのだろうか。それなしには、脳機能の本当の理解が得られないからである。

まず第1に、ヒトのように考えるコンピュー



タやヒトのように動くロボットを作るためには、つまり脳を創るためには、計算論的な理解が必須であることは明らかであろう。さらには実は、脳を守るためにもシステムレベルと計算論的な理解は必須なのである。ほとんどの病気は複雑な現象であり、1個の遺伝子や1個のリセプターがあるとかないとかいった単純な原因で起きることはまれである。たとえば、黒田らはプルキンエ細胞のシナプス可塑性、長期減弱の細胞内シグナル伝達経路のシミュレーション実験に成功したが、可塑性を理解するためには大規模で複雑な非線形の力学系の振る舞いの定量的な理解が必要であった<sup>(1)</sup>。子供の発達や教育に脳神経科学の成果を応用しようとする「脳を育む」「脳を癒す」などのためにも機能の理解、つまり定量的なダイナミクスの理解は必須である。しかし、なによりも、脳の機能が本当にわかったと納得できるのは、機能とそれを支える機構、さらに言えば情報処理が明らかにされたときであり、これにこそ、計算論的なアプローチが必須なのである。一般の人が自分の心の働きを科学的に知りたいと思うときに、それは単にあなたの脳のこの部分が働いているんですよとか、この遺伝子が大切ですよとか、このニューロンの興奮はこの心の動きと相関がありますというレベルではないはずである。なぜこういうふうに感じるかというメカニズムを知りたいはずである。これを導くためには、どうしてもシステム神経科学と計算論的神経科学の共同が必要なのである。

### ヒト知性の神経科学の可能性

ヒトに特有な知性については、計算論的な理解はできないのだろうか。ヒトについては、単一ニューロンの電気記録をとる手法は使えない。神経回路の同定も基本的にはできない。破壊実験等もちろんできないから、損傷が限局し完全であることはまれな患者のデータに頼らざるを得ない。計算論的研究の必要条件となるシステム神経科学のほとんどすべての手法が直接的には使えないのである。

それでも、ヒトに言語活動、意識に関わる認知活動などを行わせて非侵襲的に脳活動を計測して

場所に関する粗い知識を得ることはできる。脳への磁気刺激によって、可逆的な破壊実験に近いことや、脳部位間の機能的結合に関するある程度の知識を得ることはできる。さらに、脳活動と認知機能に関する適当な説明変数、あるいは脳活動と認知機能に関する任意の現象論的理論のある側面との相関を調べることもできる。しかしこういった研究戦略が脳機能の理解という意味では無力であり、また、厳密科学としては信用できないことは、Georgopolousの例からもおわかりいただけるはずである。第1次運動野の例では、ニューロンからの電気記録がなされ、解剖学的な結合は粗くはわかっていたし、計算問題はなんと言ってみてもわかりやすい運動制御だったのである。例えばヒトの言語活動であれば、ニューロンの電気記録はもちろんなく、神経回路もわからず、ましてや言語コミュニケーションを適切に説明する計算理論もない状況で研究しなくてはいけない。

ではヒト特有の認知機能の神経機構を理解する見込みがまったくないかということ、そうではない。3つの突破口があり得る。第1は今後に期待される非侵襲脳活動計測方法の革命的な進歩である。例えばfMRI(機能的核磁気共鳴画像法)と脳磁計を組み合わせて、時間と空間の分解能をそれぞれの長所にまで引き上げる方法である。しかし、これでも電気生理学や解剖学で得られる情報の詳細さと質には決して到達しないであろう。

第2はシステム神経科学で認知機能研究の実験動物として使われるマカクサルとヒトの脳の連続性である。つい10年前までは、サルの脳にはブローカ野がないから、サルは言葉を喋れないのであるというまことしやかな説が大手を振ってまかりとおっていた。ブロードマンがヒトとサルの大脳皮質の領野に番号付けをするときに違う基準を用いたためにこのような誤解が生じてしまったのである。最近、Petridesら<sup>(10)</sup>は、サルとヒトで同じ細胞構築学的基準を用いると、サルの脳にもブロードマンの44,45野が、ほぼ対応する位置(ただし弓状溝の中)にあることを示した(図1)。di Pellegrinoら<sup>(11)</sup>、Rizzolattiら<sup>(12)</sup>は、運動前野の下吻側部F5野(弓状溝後壁)で、運動の実行と観測の両方で興奮するミラーニューロンを発



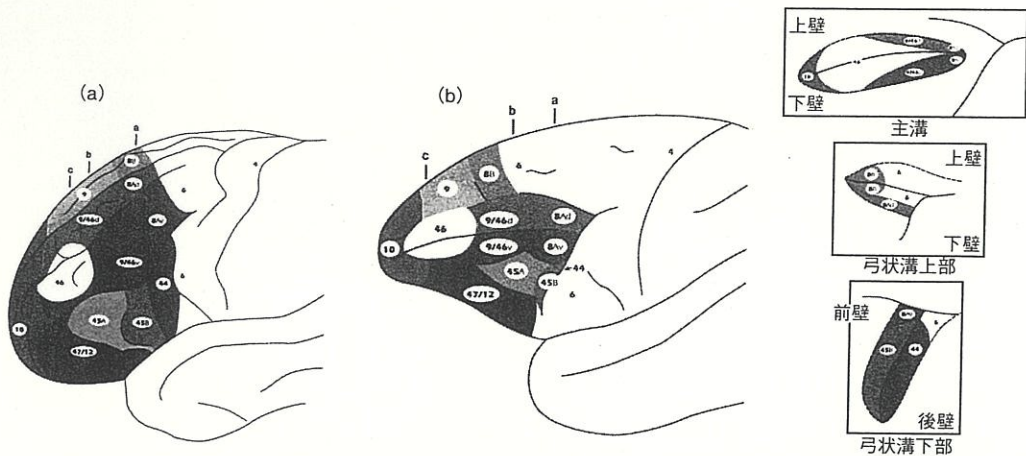


図1——ヒト(a)とマカクサル(b)の脳皮質前頭葉の細胞構築学に基づく領野への分割<sup>(10)</sup>。ミラーニューロンはヒトのプロローカ野の44野に対応する弓状溝後壁に存在する。三つの挿入図は主溝と弓状溝の中を開いて示してある。

見した。マカクサルのF5野はヒト脳のプロローカ野のうち44野に相同である。ミラーニューロンはサル自身が食物を机に置いたり、取ったり、他人の手から食物を把持したり、紙を引き裂いたりなどの意味のある手の運動を行ったときだけでなく、実験者あるいは他のサルが同じ運動を行ってそれを視覚的に観察したときにも興奮する。この生理学的データは明らかに同じハードウェアが運動の実行と観測で使われていること、そしておそらく同じ表現が使われていることを示唆している<sup>(13)</sup>。

Rizzolattiら<sup>(12)</sup>、Graftonら<sup>(14)</sup>は上記のサルの生理実験に対応するヒトPET計測実験を行った。まず把持運動の観察と実行を比較して、サルの生理実験から示唆される上側頭溝およびF5野に相当する脳部位が活動するか否かが検証された。その結果ブロードマン21野とブロードマン45野が把持運動の観察で活動し、運動の想像でブロードマン44野が活動した。ヒトでもプロローカ野は言語機能のためだけに働くのではなく、運動の観測や想像というより包括的な機能のための脳部位なのである。小脳外側半球とプロローカ野は大脳小脳連関によって双方向に結合されていることが解剖学的に知られているが、このループが運動の観測や想像において活動している。プロローカ野がより複雑な構文を理解するとき、より激しく活動すること<sup>(15)(16)</sup>、またサインランゲージの理解でも活動すること<sup>(17)</sup>をも考え合わせると次のよう

な外側言語システムの連続的な進化の仮説が導かれる。つまり、プロローカ野と小脳半球腹外側部が作るヒトの外側言語処理系は、顔の表情や手のジェスチャーを認識するための、サルにも存在するより古いシステムから進化してきた。

第3の突破口は、これまでの5回の連載で紹介してきたヒト知性の計算理論の進歩である。これまでの連載で紹介してきた計算理論(第2回「小脳の内部モデル」、第3回「大脳皮質、小脳、大脳基底核の異なる学習方式」、第4回「モザイク」、第5回「強化学習モザイク、階層モザイク」)に基づいて、コミュニケーションの計算モデルを定式化しよう。

#### モザイクに基づくコミュニケーションのモデル

プロローカ野、ウェルニッケ野、小脳外側部、大脳基底核前部が、運動パターンの認知・編集・計画・実行に用いる神経計算原理が、言語か非言語かを問わずコミュニケーションの基礎となる。この過程で重要なのは、運動系列を分節化し、各単位を認知し、自由に入れ替え、特定の行動目標のために運動単位の順序を計画し、かつそれを実行する能力である。このためには、運動系列の分節化に対応して、単位となる運動要素ごとに神経系にモジュールが学習で獲得されなければならない。複雑な運動系列をくり返し実行するあいだに、学習によってモジュールがモザイクの中に自己組織



的に形成されていくのであった(連載第4回)。これは、運動系列を教師なしで、分節化していることになる。また階層モザイク構造の上の階層のモジュールは、感覚フィードバックや運動指令とは直接関係のない高次の概念のシンボルを表現することになる(連載第5回)。したがって階層モザイクは、シンボル接地問題を解き、コミュニケーションのためのモジュール構造を作る能力がある。

コミュニケーションの過程そのものは、他者が示した運動パターンから他者の内部状態を推定するコミュニケーション信号の認知の問題と、運動制御によってある運動パターンをつくり出しそれを他者に提示することによって他者の内部状態を制御するコミュニケーション信号による制御の問題の2つに分かれる。われわれの仮説は、言語的コミュニケーションの基礎に、非言語的コミュニケーションが存在し、両者の神経計算原理は基本的には同じということである。言語コミュニケーションの認知は、連続的な信号からシンボルへのパターン認識であり、生成はシンボルから連続的信号への変換であるというのは、古典的な見方で、一見わかりやすいが本質的に誤っている。パターン認識の工学的研究からわかったように、言語音声にしる、文字にしる、あるシンボルを表現するコミュニケーション信号に不変な特徴があって、そのおかげで簡単に分類できるわけではない。外部記憶装置として使われる現代的なシンボルは人類の歴史で言っても数千年の歴史しかないわけで、シンボル表象に完全に依存する上のようなコミュニケーションの理論は生物学的に誤っている。私たちのコミュニケーションに関する新しい理論は、コミュニケーションを他者の脳の内部状態に対する運動制御と認知であると見なすことから始まる。この枠組みで、計算論的神経科学が築いてきた、運動制御と認知に関する理論や概念が敷衍できるのである。

モザイクに基づく、コミュニケーションの計算理論は、次の5つの点で、コミュニケーションを感覚運動制御から統一的に理解する魅力的な枠組みを与える。(1)コミュニケーションの神経計算原理と、感覚運動制御の神経計算原理の間に不

連続な飛躍を仮定しない。(2)コミュニケーションを他者の隠れた内部状態の推定とその制御という、計算問題としてすっきりと定式化する。(3)言語コミュニケーションで中心課題となるシンボルについて、ニューロンもしくは神経系の状態とシンボルを1対1にあらかじめ天下りに対応させるのではなく、感覚運動制御を通して、シンボルが神経系のハードウェアもしくはそのアトラクターとして自己組織される過程までも説明する。(4)モザイクの階層化を通して、コミュニケーションの階層性、さらには文法などの問題を計算論的に統一的に扱える。(5)コミュニケーションの最も基本的な機能であると考えられる見まねの能力を、モデルの中で自然に取り扱える。

### コミュニケーションの計算問題としての難しさ

コミュニケーションの計算理論を定式化しよう。コミュニケーションは少なくとも2つの逆方向の情報処理過程を含んでいる。つまり、他者の隠れた内部状態の推定とその制御である。第1の過程は、隠れた変数の推定という意味で難しい。第2の過程は、ある未知のプロセスの制御という意味で難しい。しかしながら、われわれはすでに感覚運動制御の研究でこのような困難を経験し、それを解決するためには何が必要かを知っている。前者は、視覚の計算論で画像情報から3次元世界の様子を推定するときを生じた困難に似ている。後者は、運動制御において、未知の制御対象をフィードフォワード制御するときを生じた困難に似ている。視覚では、画像が作られる過程(光学)のモデルと、3次元世界に関する事前知識を組み合わせることによって、不良設定問題が解かれた<sup>(18)</sup>。運動制御においては、制御対象の内部モデルが運動学習によって獲得され、制御に使われた<sup>(9)</sup>。コミュニケーションの推定と、制御の両方で、同様に他者の内部構造のモデルが必要となる。

図2は、エージェントBが、エージェントAのつくり出したコミュニケーション信号(運動パターン)を認知する過程を、復号化として捉えたブロック図である。つまり、認知の過程は、エージェントBが、エージェントAが発した運動パ



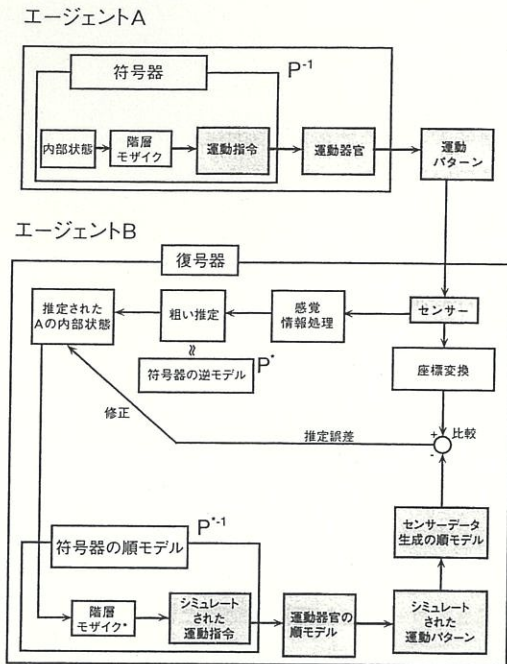


図2—エージェントBがエージェントAの運動パターンを認知する過程。エージェントBは逆モデルを用いてエージェントAの運動パターンからその内部状態を推定する。続いて推定した内部状態から順モデルを用いてAの運動パターンをシミュレートし実際の運動パターンとの誤差を学習信号として次第に推定の精度を高める。

ターンからエージェントAの内部状態を推定(復号化)するプロセスである。エージェントAは、その内部状態から、運動指令をつくり出す符号器を持っている。その特性を $P^{-1}$ で表わす。エージェントBは、この符号器の順モデルを持っている。その特性は、完全に同一ではないから、 $P^{*-1}$ で表わす。エージェントBの脳内には、エージェントAの内部状態の表現があり、これを符号器の順モデルに入力して、予想される運動パターンをつくり出す。この予測と、実際に観測した運動パターンを比較して、内部状態の推定をより精密にする。この過程がモザイクの枠組みで、効率的に行えることは、連載第5回の2の図5と図9などですでに解説した。この計算モデルの最も本質的な仮定は、コミュニケーションの認知の過程に、コミュニケーション信号の生成の過程を使っていることである。その意味で、認知科学的に言えば、コミュニケーションのシミュレーション説に基づいている<sup>(19)</sup>。

図3は、エージェントBが、エージェントA

の内部状態を、コミュニケーション信号(運動パターン)を生成して見せることによって、変更(制御)する過程を、符号化として捉えたブロック図である。つまり、コミュニケーションの信号伝達の過程は、エージェントBが、自ら発した運動パターンによってエージェントAの内部状態を制御するプロセスである。エージェントAは、図2で、説明した復号器によって、エージェントBの発した運動パターンから、Bの内部状態を推定し、それによって自身の内部状態を変更する。この復号器から、内部状態変更までの過程の特性を $Q$ で表わす。エージェントBの運動指令をつくり出す符号器の特性は、上記の過程の近似的な逆であれば、効率的な制御が行えるので、それを $Q^{*-1}$ で表わす。エージェントBの脳内には、エージェントAの望ましい内部状態の表現があり、これを符号器に入力して、運動パターンをつくり出す。この一連の過程は、制御対象が単なる筋骨格系や道具ではなく、知性をもった他人であるという違いはあるが、運動指令を発することによって、世界の状態を変更するという意味では、相変わらず、運動制御なのである。

以上、コミュニケーションと一般的な運動制御、認知の問題の共通性に着目して説明したが、コミュニケーションは計算問題としてずっと難しくなっている。コミュニケーションの対象が他者の脳の内部状態という直接観測できない、隠れた変数だからである。この場合にも他者の脳に関する良い内部モデルがあれば、このような計算はある程度行える。しかし、他者の脳は、運動制御対象や道具などと比べてずっと複雑、大規模で予測の難しいシステムであり、したがって内部モデルを作ることが難しい。さらに、他人の脳という、入力と出力、内部変数が直接観測できないシステムの内部モデルを獲得しなければいけない。したがって、他者の脳の内部状態の推定、および他者という未知の制御対象の制御は、特別な拘束条件なしには解けない。唯一の理論的可能性は脳が互いに似通っていることに基づく方法である。ヒトに共通な遺伝情報により、階層モザイクの基本構造が同一であることを利用して、詳細な構造も学習と教育などの社会的相互作用によって、相似形に収



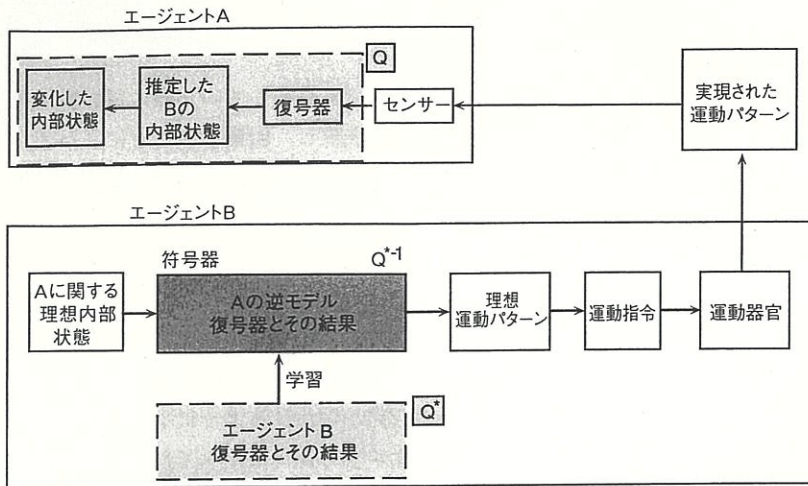


図3—エージェント B がエージェント A の信念に作用する過程。エージェント B はエージェント A 内に作り出した理想内部状態からエージェント A の逆モデルを用いて運動パターンを生成する。これを観測したエージェント A は復号器を用いて B の内部状態を推定し自分の信念を変更する。

束させるのである。次にこれが可能であることを示す階層モザイクに基づくシミュレーションを説明する。

### 心の理論の計算モデル

ヒトの知性がそれ以外の動物の知性と異なる、ハードウェアとしての最大の違いは、他者の脳の良い内部モデルを学習と教育によって獲得し、それをコミュニケーションに使用できるほどの、複雑さを持つ点にある。ここまで説明したコミュニケーションの計算理論でも、他者の脳の内部モデルが本質的な役割を果たしたし、またそれを獲得することは感覚運動制御の内部モデルに比べて格段に難しいことも指摘した。いよいよ、図4を用いて、2人のエージェントが、遺伝的に似通った脳の内部構造を持ち、同じ環境の中で他者の行動を観察することで、お互いの内部モデルを学習で獲得できることを説明しよう。この過程では、連載第5回の2で説明した、階層モザイクによる他者の内部状態の推定と見まねが再び本質的な役割を果たす。

2人のエージェントが似通った遺伝情報を種として、それぞれ共通の世界に働きかけ、すでに似ているが異なる脳の内部構造(階層モザイク)をそれぞれ獲得しているとする。ここでの問題は、自己の脳内に他者の階層モザイクの内部モデル(とりあえず順モデルを考える)を獲得することである。まずそれぞれの階層モザイクは内部状態と

して責任信号等を持ち、感覚入力を受け、運動指令を出力する。この3種類の状態変数がすべて観測できるなら、内部モデルの学習による獲得は特別に難しくはない。ところが、感覚入力も完全には推定できないし、運動指令の推測も大変困難である。ましてや責任信号などの内部状態は直接計測できない。そこで、エージェント A は、エージェント B の運動を視覚的に観察することによって、その運動指令と内部状態を推定する。その計算は連載第5回の2の図5と図9などです。すでに解説した階層モザイクの見まね学習時の運動指令と内部状態の推定と同じである。こうして推定した、運動指令と内部状態と観測した感覚フィードバックの組を教師信号として、エージェント A が自己の脳内に、エージェント B の階層モザイクの内部モデルを作るのである。春野らはこれが可能であることをシミュレーションによって確認している。

われわれが他者の内部モデルと呼ぶ計算論的実体は、発達心理学や霊長類研究で話題になっている心の理論に対応する<sup>(20)</sup>。ヒトでも4歳になってようやく獲得し、チンパンジーではその存在が論争的になっている。心の理論は、計算論的に考えても、またハードウェアとして考えても、確かにその学習による獲得は難しいのである。ここにヒトの知性の秘密が潜む可能性は高い。



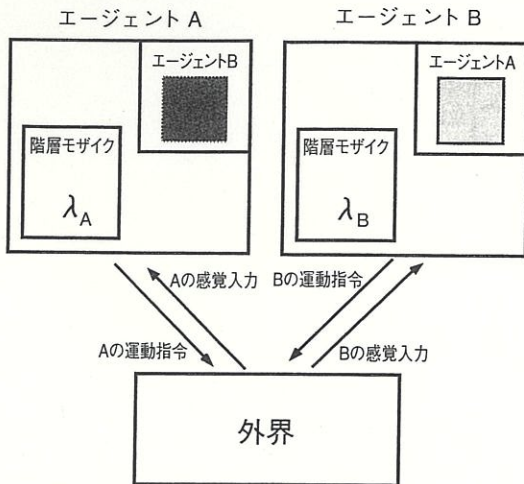


図4——心の理論の計算モデル。エージェントA, Bはそれぞれ相手の運動指令, 感覚入力, 責任信号 $\lambda$ を直接観測することはできない。そこで各エージェントは相手の運動パターンを観測し階層モザイクによりその運動を生成するための運動指令と責任信号を推定する。順モデルにより得られた予測運動パターンと実際の運動パターンの誤差を用いて学習を行うことで次第に相手の意図を理解できるようになる。

## 言語コミュニケーションから 非言語コミュニケーションへ

考古学的なデータから、声道の断面形状が現代的な形を取り始めたのは30万年前の古い型のホモサピエンスの発生以降であると考えられている。石器や絵画など遺物に見られる複雑化の度合を指針として、言語は人類の先史時代の末期3万5000年程度前に急激に発生したと考える研究者もいる。このような考えに基づいて、Donald<sup>(21)</sup>は、ヒト文化と認知の進化に3つの段階があると主張している。模倣の技能(ミメシス)、言語、そして外部シンボルである。ミメシスはホモエレクトゥスが保有していたと考えられる認知能力である。模倣行動は認知された事象を行動に翻訳するために全身をコミュニケーションの道具として使う才能である。ミメシスは、身体を環境の中でとらえる知覚モデルによって導かれていて、運動を随意的にリハーサルして洗練させる記憶システムに基づいている。したがってリハーサルの産物を格納することも読み出すこともできる。ミメシスは、意識的な制御の下に、任意の身体の随意運動を止めたり、再生したり、あるいは編集したりす

ることを可能にする抽象的な「モデルのモデル」に基づいている。

ミメシスはすでに存在していた類人猿の表現をリハーサルし、洗練し、編集することによって、顔、音声、そして全身の表現の多様性を劇的に増加させた。これによって道具の製作や使用の教育が格段に進みホモエレクトゥスの洗練された石器製作を可能にした。またミメシスが可能にしたずっと豊かなコミュニケーションによって社会的な協同作業が可能となって、季節ごとのキャンプや、長い距離の狩猟が行われるようになった。

言語の進化が脳の進化に比べて最近のできごとであるとすれば、ヒト言語の起源をすべて脳の中に求めるのは大きな誤りである。言語器官や文法遺伝子などの用語・概念はまさにそのような誤りの象徴である。次節で説明するように、言語の発生は、一般文法にしろ、セマンティックスにせよ、集団遺伝学や数理生態学的手法を駆使して、コミュニケーションの効率化によるフィットネスの上昇を契機として、複数の脳の間ダイナミクスとして生じた現象として説明されるべきである<sup>(22)~(25)</sup>。

## ヒト知性の神経科学的方法論

これまでの議論をまとめてヒト知性の神経科学の今後の方法論を提起しよう。ヒト知性、特にコミュニケーション・言語のメカニズムを神経科学的に解明するためには、(1)言語の発生と進化への数理生態学的アプローチ、(2)行動とコミュニケーションの計算理論、(3)脳の計算モジュール構造の解明、(4)局所回路、細胞、物質レベルでの研究を、それぞれ同時並行的に、進めることが必要である。

### (1)数理生態学的アプローチ

近年 Nowak らは、事物に対して任意に記号を割り振ることのできる個体の集団を考え、うまくコミュニケーションを成し遂げた個体ほど、高い確率で同じ事物-記号の関係を使う子孫を残すという数理生態学的なモデルのもとで、集団の中で共通の事物-記号の対応が使われるようになる



条件を求めた。また彼らは同様な淘汰の仮定のもとで、集団に固有の文法構造が形成される条件、さらに個別の文法を獲得するメカニズムとしての普遍文法が、進化の過程で形成され安定に存続し得るための条件を求めている<sup>(22)-(25)</sup>。このような研究は、事物の分節化や文法処理の実際など、言語処理の具体的な問題を捨象してはいるものの、言語コミュニケーション研究全体を、生物学的な枠組みで理解する上での基礎を与えるものである。

実世界で行動するロボット集団の進化・適応の過程で、事物の分節化、記号との対応、文法構造がいかにか形成され得るかに関する実験、あるいはシミュレーションによる研究も最近さかんである<sup>(26)</sup>。これらの実験的取組みは、上記の理論的アプローチとともに、言語の発生と進化の生態学的基盤を明らかにし、その神話を解消する役割を果たすと考えられる。日本のロボット研究コミュニティにかかる期待も大きい。

## (2) 行動とコミュニケーションの計算理論

感覚と行動の分節化、モジュールの自己組織化、隠れ状態の推定、それに基づく系列運動制御など、感覚運動制御にねざした形でのコミュニケーションの計算理論の確立が必要である。

上記モザイクアーキテクチャは、そのひとつのひな形を与えるものである。しかし、高度に並列的、階層的な学習系が、縮退や干渉などに容易に陥りやすいという問題を回避しつつ、いかに効率よく機能し得るかを探る課題はまだ多数残されている。学習機械や音声・言語処理の理論的、応用的研究は、その上でいくつかの手がかりを与えてくれる可能性がある。これまでにも、独立成分解析、カルマンフィルタ、強化学習など、元々は工学の分野で育った理論が、実は脳のモデルとして有用であったという例はいくつもある。このような理論分野での最新の展開を理解して、それらを脳の情報処理のモデルとして解釈し拡張する柔軟な視点を持つことが重要である。

## (3) 脳の計算モジュール構造の解明

コミュニケーションにおいては、他者の認知機構や行動則の内部モデルと、その評価を与える報

酬モデル、さらにそれらの予測を保持する作業記憶のメカニズムが必要である。これらの計算要素が、脳のどの部位によって担われているかを明らかにすることは、単にfMRIなどの脳活動計測だけでは難しい。ほとんどの場合、程度の差こそあれ、それらはほぼ同時に活動し、協調的に働いているからである。

その解明のためには、コミュニケーションに必要な計算要素に的を絞った行動に対するfMRI計測を行い、さらに脳磁計のようなより時間分解能の高い手法を組み合わせることが必要であろう。さらに脳の異なる部位が担う計算処理の違いや、解剖学的構造、サルなどでの実験結果と組み合わせることによって、計算理論が予測するモジュール構造の検証とその拡張が可能になる<sup>(27)(28)</sup>。

## (4) 神経回路、細胞、物質レベルの研究

ある細胞の活動や遺伝子の破壊などと行動の相関を見るだけでは脳のメカニズムの解明にはつながらないということは、繰り返し述べた。上記のような、生態学的アプローチ、計算理論モデル、その脳回路での実現に関する知見と照らし合わせた上で、初めて個々の神経細胞の活動や、その細胞、分子、遺伝子レベルでの知見を、コミュニケーションのメカニズムにつなげて理解することができる。

ヒト知性の神経科学的理解には、これらのアプローチを、相補的かつ有機的に組み合わせることが必要である。特にこれまでの神経科学との大きな違いは、単に個々の個体の脳だけを見るのではなく、他の個体の脳との相互作用、さらに脳とその外にある文化的、社会的要素との同時進化をも考慮する必要があるという点にある。

## 望まなければ得られない

ヒト知性の機能、情報処理を自然科学として理解することは難しい。しかし、希望がないわけではない。最終回を要約すれば、4つのレベルの研究を、計算理論を軸として連結させることでこれが可能になる。

第1のレベルでは、動物を対象にしたシステ



ム神経科学で、サルにも備わっているが、ヒト知性の重要な構成要素となる、内部モデル、強化学習、作業記憶などの認知機能について神経生理学や解剖学のデータを得る。第2のレベルでは、サルとヒトで共通のこのような認知機能についての計算論的研究を展開し、ヒトでの非侵襲脳活動計測からも検証する。第3のレベルでは、第2のレベルと連続性を保ちながら、ヒト特有の知性の計算理論を探り、非侵襲脳活動計測で検証していく。第4のレベルでは、コミュニケーションと社会の計算理論を、脳の計算理論、現実的な環境のモデル、そして数理生態学、社会ロボティクスなどの手法を組み合わせて構築する。生成文法研究そのものの価値をおとしめる意図はまったくないが、連載第1回で批判した4つの仮定に基づくと、研究は第3のレベルだけに閉じてしまい、生物学の成果を利用しないことになって、自然科学として成立しないと危惧する。また、脳の場所や、物質、相関に関する研究は、信頼できる自然科学の知識として重要だが、そこだけに満足すれば、ヒトの知性の機能を理解することにつながらない。

ヒトの知性を自然科学として理解しようという目標がそもそもだいそれていて無理なのだときらめる(両方からの)立場もあるだろう。それは現段階で効率の良い研究プログラムを遂行する上では合理的かもしれない。しかし、望まなければ得られないのである。この6回の連載が、21世紀の新しい学問分野が生まれるための議論を刺激することができたなら、著者らの大きな喜びである。

#### 文献

- (1) S. Kuroda et al.: *J. Neurosci.*, **21**, 5693(2001)
- (2) A. P. Georgopoulos et al.: *Science*, **233**, 1416

- (1986)
- (3) S. Kakei et al.: *Science*, **285**, 2136(1999)
- (4) E. Todorov: *Nature Neurosci.*, **3**, 391(2000)
- (5) S. H. Scott et al.: *Nature*, **413**, 161(2001)
- (6) K. Kawano: *Curr. Opin. Neurobiol.*, **9**, 467(1999)
- (7) A. Takemura et al.: *Journal of Neurophysiol.*, **86**, 1750(2001)
- (8) M. Shidara et al.: *Nature*, **365**, 50(1993).
- (9) M. Kawato: *Curr. Opin. Neurobiol.*, **9**, 718(1999)
- (10) M. Petrides & D. N. Pandya: In *Handbook of Neuropsychology* 9, Elsevier Science B. V.(1994) pp. 17~58
- (11) G. di Pellegrino et al.: *Exp. Brain Res.*, **91**, 176(1992)
- (12) G. Rizzolatti et al.: *Cog. Brain Res.*, **3**, 131(1996)
- (13) G. Rizzolatti & M. A. Arbib: *Trends in Neurosci.*, **21**, 188(1998)
- (14) S. T. Grafton et al.: *Exp. Brain Res.*, **112**, 103(1996)
- (15) K. Stromswold et al.: *Brain and Language*, **52**, 452(1996)
- (16) M. A. Just et al.: *Science*, **274**, 114(1996)
- (17) D. Bavelier et al.: *Neuron*, **21**, 275(1998)
- (18) M. Kawato et al.: *Network: Computation in Neural systems*, **4**, 415(1993)
- (19) V. Gallese & A. Goldman: *Trends in Cognitive Sciences*, **2**, 493(1998)
- (20) C. D. Frith & U. Frith: *Science*, **287**, 1692(1999)
- (21) M. Donald: *Origins of the Modern Mind*, Harvard University Press, Cambridge(1991).
- (22) M. A. Nowak et al.: *Nature*, **404**, 495(2000)
- (23) M. A. Nowak et al.: *Science*, **291**, 114(2001)
- (24) M. A. Nowak & N. L. Komarova: *Trends in Cognitive Sciences*, **15**, 288(2001)
- (25) T. Hashimoto & T. Ikegami: *Biosystems*, **38**, 1(1996)
- (26) L. Steels: *Artificial Intelligence*, **103**, 1(1998)
- (27) K. Doya: *Current Opinion in Neurobiology*, **10**, 1732(2000)
- (28) K. Doya: In *Affective Minds*, Elsevier Science B. V.(2000)pp. 101~104