

科学

Science Journal *KAGAKU*

VOL.70 2000 別刷

岩波書店

■ヒト知性の計算神経科学<第3回>

小脳，大脳基底核，大脳皮質の 機能分化と統合

銅谷賢治 川人光男 春野雅彦

脳の高次機能は，大脳新皮質によって担われるというのがこれまでの固定観念であった。しかし，近年の実験データは，小脳や大脳基底核が，大脳皮質とのループ回路によって，運動制御だけでなく多様な認知機構にかかわることを示している。小脳，大脳基底核，大脳皮質がそれぞれ，“教師あり学習”“強化学習”“教師なし学習”という学習のアルゴリズムに特化した組織だという仮説のもと，それら学習モジュールをつなぐ脳全体の回路で，行動結果の予測とその評価に基づく行動とコミュニケーションがいかにか可能になるのか，その基本的枠組みを示そう。

人間の高度な知能は大きく発達した脳の機能によるものだ，というのは常識であるが，“脳”という場合，多くの人は曲がりくねったしわをもつ大脳新皮質を思い浮かべるのではないだろうか。連載第2回では，大脳の下に半分かくれている小脳が，運動制御だけでなく，“内部モデルの獲得”という機能によって，想像，思考や言語など，広く高次認知機能に関与することを述べた。今回はさらに，大脳半球の内側にかくされた“大脳基底核”にスポットを当て，小脳，大脳基底核，大脳皮質をつなぐ脳のグローバルな(脳全体としての)回路で(図1)，どのような機能分担と統合がおこなわれ，行動学習やコミュニケーションが実現されるかについて検討しよう⁽¹⁾。

大脳基底核は，パーキンソン病やハンチントン舞踏病など，その病変によって顕著な運動障害がおこることから，運動制御に関与することが想定されていたが，その正常時における機能は，長い

間謎に包まれていた。近年，大脳基底核で強く作用するドーパミンと呼ばれる神経伝達物質系の研究から，大脳基底核は，将来得られるであろう報酬の予測をもとに目的指向的に行動を組み立てる上で，主要な役割を担っていることが明らかになってきた⁽²⁾。

小脳の外側部と大脳基底核はともに，大脳皮質の主に前頭連合野との間で並列的なループ回路を構成する。最近，特殊なウイルスを使い神経線維をトレーシングする技術によって，小脳と大脳基底核の出力は，一次運動野や運動前野のような運動に直接関連する領野だけでなく，高次の認知機構にかかわる前頭前野にも投射していることが明らかにされた(図2)⁽³⁾⁽⁴⁾。また，最近のPET(陽電子放射断層撮影)やfMRI(機能的磁気共鳴画像法)などを用いた多くの実験データは，小脳と大脳基底核が運動以外の課題にも関与していることを示している。これらの事実は，小脳と大脳基底核の機能を，運動制御という枠組みだけで考えていた，従来のパラダイムの転換を迫るものである。

それに代わる新たな視点は，小脳，大脳基底核，大脳皮質が，特定の機能の実現のためではなく，それぞれ独自の学習アルゴリズムのために専門化

Kenji DOYA ATRインター(国際電気通信基礎技術研究所)，科学技術振興事業団戦略基礎研究脳を創る，Mitsuo KAWATO ATR人間情報通信研究所，科学技術振興事業団創造科学技術推進事業川人学習動態脳プロジェクト，Masahiko HARUNO ATRインター(国際電気通信基礎技術研究所)

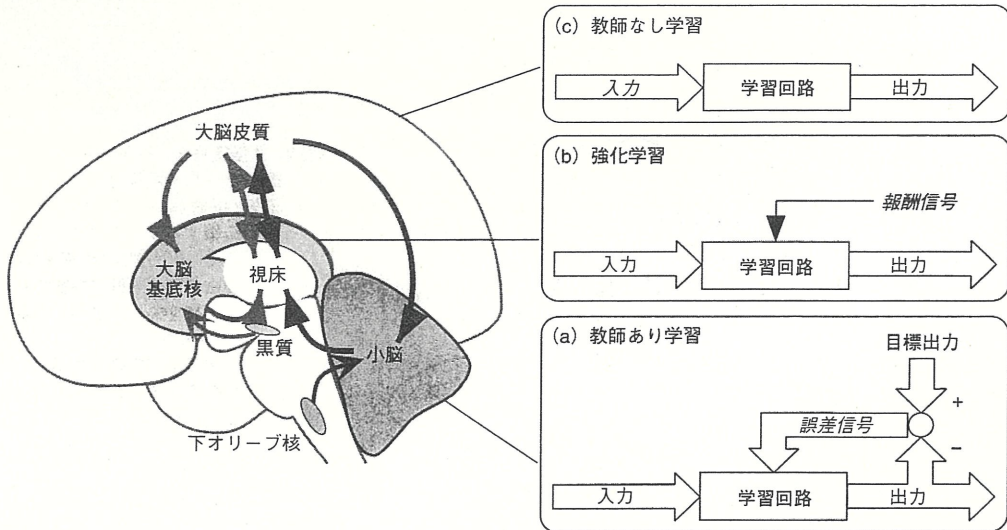


図1 小脳、大脳基底核と大脳皮質との間のループ回路。大脳皮質は、大脳基底核と小脳へ出力を送る。大脳基底核と小脳はそれぞれ、視床を経て大脳の前頭葉を中心とした領域に向けて出力を送り返し、並列的なループ構造をなす。小脳、大脳基底核と大脳皮質は、このループによって連携して働きながら、それぞれ以下のように、異なる三つの学習のパラダイムに専門化していると考えられる。(a)小脳: 出力の誤差ベクトルによる教師あり学習。(b)大脳基底核: スカラー値の報酬信号による強化学習。(c)大脳皮質: 入力信号の統計的構造による教師なし学習。

した組織であるという見方である⁽¹⁾⁽⁵⁾。

以下に述べるように、小脳は、“教師あり学習”によって獲得した外界の内部モデルを提供し、行動の結果の“予測”をおこなう。大脳基底核は、“強化学習”によって“報酬”の予測をおこない、ある行動の結果に評価を与える。大脳皮質は、“教師なし学習”によって、外界の状態や行動出力の必要かつ十分な内部表現を提供する。

つまり、状態の“予測”と“報酬”という二つの重要な計算要素が小脳と大脳基底核によって提供され、それらが小脳と大脳皮質を受け皿として組み合わせられ、複雑な行動の計画や実行が実現されることが可能になる。従来の大脳皮質中心の見方は、その一断面を捉えているにすぎない。

学習の三つの枠組み

人間や動物の行動は学習によって変化する。つまり、感覚入力と同じでも、それに対する行動出力は、その環境に適したものに次第に変化する。脳の回路の各部分部分も、また生物だけでなく、機械やロボット、プログラムも、経験に応じて入出力関係を変化させることができ、学習が可能で

ある。しかし、ただ変化するというだけでは学習だか暴走だかわからない。そこで学習には、変化に対する何らかの評価基準が必要となる。図 1a, b, c は、このような学習の理論的研究で扱われる、三つの主な学習のパラダイムを示したものである⁽¹⁾⁽⁶⁾。

“教師あり学習”では、入力に対する正しい出力が教師信号として与えられ、実際に出した出力との誤差信号がゼロになるように学習がおこなわれる。

いっぽう“強化学習”では、どういう出力を出せばよいかという具体的な教師信号は与えられないが、実際に出してみた出力がどれくらいよかったかを示す“報酬信号”が与えられ、これをなるべく大きくするよう探索的に学習がおこなわれる。

さらに“教師なし学習”では、どういう出力を出すべきかはまったく教えられない。そこで入力信号自体の統計的分布をもとに似たもの同士をグループ分けしたり、信号を独立な成分に分解したりという形で学習がおこなわれる。

脳の神経回路の構造、シナプス可塑性、ニューロン活動の記録の多くのデータから、小脳は、下オリーブ核から登上線維によって送られる誤差信

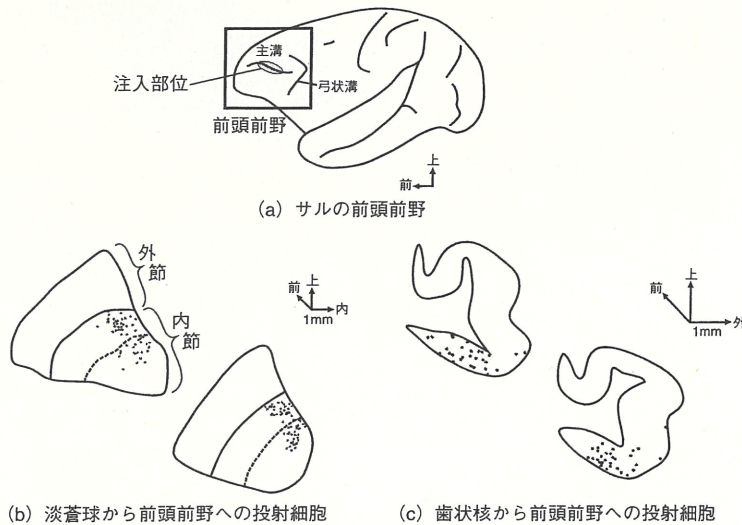


図2 大脳基底核と小脳から、大脳皮質前頭前野への神経連絡。(a)サルの前頭前野(46野)に、神経線維をさかのぼって伝搬するウイルスを注入。(b)大脳基底核の出力部位(淡蒼球)の切片。視床経由で前頭前野に投射する細胞がウイルスに感染し、黒い点で示されている。(c)小脳の出力部位(齒状核)の切片。同じく視床経由で前頭前野に投射する細胞が示されている。文献⁽²⁾から引用。ほかに、大脳基底核と小脳の異なる部位が、運動前野、補足連行野、一次運動野、頭頂葉、側頭葉などへ投射することが明らかにされている⁽⁴⁾。

号をもとにした教師あり学習(小脳については前回参照)、大脳基底核は、中脳の黒質ドーパミン細胞から送られる報酬信号をもとにした強化学習、大脳皮質は、その相互結合回路のダイナミクスによる教師なし学習に、それぞれ専門化した組織であることが示唆されている⁽¹⁾⁽⁵⁾⁽⁶⁾。

動物の行動学習は、全体としては、生存と繁殖にかかわる水、餌、交尾行動などの報酬を得るための強化学習のプロセスと捉えることができる。しかし、全体としては強化学習の課題でも、その一部に教師あり学習の要素が使われたり、教師あり学習を効率よくおこなうために、あらかじめ教師なし学習で情報を抽出したりということはしばしばおこなわれ、これら三つの学習の枠組みは、一般にさまざまな組合せで用いられる。

小脳の教師あり学習

小脳の学習が、プルキンエ細胞への入力シナプスが、その入力信号と、登上線維によって送られる誤差信号との相関にもとづき変化する長期抑圧(LTD)によって、出力誤差をゼロにするような教師あり学習であることは、前回詳しく述べた。小脳には、このような教師あり学習をおこなう要素回路(モジュール)が多数並列に備わっているわけだが、これらは、行動学習上どのような役に立つのだろうか？

ひとつは、身体や環境のさまざまな特性を近似する内部モデルを提供することである。例えば、現在の身体の状態と、運動指令信号のコピーを入力として、実際の運動の結果を教師信号として学習をおこなえば、身体運動の“順モデル”が学習できる。これは以下に述べるように、行動の計画と学習にとって非常に重要な要素である。

また、例えば眼球運動の場合、視線と視覚刺激とのずれによって与えられる視覚の誤差信号から、眼球を動かす筋肉の配置をもとに、運動指令の誤差信号を推定することが可能である。これを使えば、身体の現在の状態と目標とすべき状態を入力として、必要な運動指令を計算する“逆モデル”が獲得されるというのが、前回解説した“フィードバック誤差学習”の原理である。

教師あり学習モジュールの使い方としてもうひとつ重要なのは、複雑な処理の過程をひとまとめにして保持しておく“カプセル化”あるいは“マクロ化”である。新しい行動の学習には、以下に述べるように、行動の探索、予測と評価など、小脳、大脳基底核、大脳皮質など脳の広範な部位を使った処理が必要である。しかし学習の結果、どの入力に対してはどの出力を出せばよいということがいつぱんわかってしまえば、後はそれを覚えておいて、必要なときに参照すればよい。これはちょうど、複雑な計算の結果を早見表やグラフのような形で取っておくと、時間と労力の節約がで

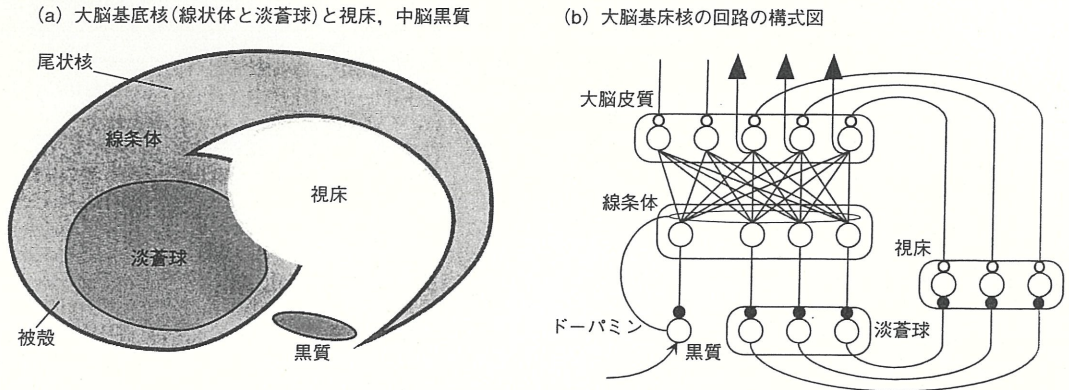


図3 大脳基底核の回路構造。(a)大脳基底核と視床, 中脳の黒質を, 脳の内側からみた模式図。入力部である線状体は, 尾状核と被殻からなり, 出力部である淡蒼球は被殻の内側にはまり込む形になっている。(b)大脳基底核の回路の模式図。線状体は, 大脳皮質の広範な領野から入力を受け, 黒質と淡蒼球に信号を送る。黒質(緻密部)には, ドーパミン細胞が集まり, 線状体に出力を送り返している。淡蒼球(および黒質の網様部)は, その一部は脳幹の運動中枢に, 多くは視床経由で大脳皮質の主に前頭葉に出力を送り, それによってループが構成される。(淡蒼球は内節と外節に分かれ, さらに視床下核を加えた間接経路と直接経路があるが, ここでは単純化のため省略した。)

きると同じである。

小脳の外側部は, 大脳皮質の広範な部位とループ状の回路を構成しており, 大脳皮質の出力を教師信号として, 脳内に実現された任意の入出力関係をコピーする形で学習することが原理的には可能である。われわれが新しい行動を学習するとき, はじめは考え迷いながら意識的におこなっていた行動が, だいに無意識的自動的に実行できるようになる背景には, このようなメカニズムの存在が考えられる。

大脳基底核の強化学習

大脳基底核は, 大脳皮質の広範な領域から入力を受け, 脳幹の眼球運動中枢などに出力を送ると同時に, 視床を経て大脳皮質に出力を返している(図3)。大脳基底核の入力部位である線状体は, ドーパミンと呼ばれる神経伝達物質がとくに多く存在することで知られている。線状体にドーパミン性の神経線維を送っているのは, 中脳の黒質と呼ばれる部分である。

ドーパミンは“快楽物質”と呼ばれることもあり, 麻薬など依存性をもつ薬物の大半は, 脳内のドーパミンの濃度を直接あるいは間接的に上昇させる働きをもつことが知られている。また, パーキンソン病の主要な原因は, 黒質のドーパミン性

の細胞が減少することであり, ドーパミンの原料になる L-DOPA などの物質を補うことで症状が改善される事例が知られている(映画‘レナードの朝’でご覧になった読者もいることだろう)。

Wolfram SCHULTZらは, サルにレバー押し課題を学習させながら, 黒質のドーパミン細胞の活動を記録した(図4)⁽⁷⁾。

サルは, ランプが点灯してその下にあるレバーを押すと1滴のジュースがもらえる。課題とは関係なく単にジュースを与えた場合や, 学習の初期にたまたまうまくレバーを押せた場合には, ドーパミン細胞はジュースがもらえた時点で反応する(図4a)。しかし, 学習が十分進み, 毎回確実にレバーを押してジュースがもらえるようになると, ドーパミン細胞はジュースに対しては反応しなくなり, かわりに, ランプが点灯した時点で応答するようになる(図4b)。これは, “これからジュースがもらえるぞ”という報酬の期待に対してドーパミン細胞が応答しているようにみえる。

では, レバーをちゃんと押して, ふつうならジュースがもらえるタイミングで, ジュースが来なかったらどうなるか。その場合, ドーパミン細胞の発火は通常以下のレベルに落ち込むことが確かめられた(図4c)⁽⁸⁾。つまり, ドーパミン細胞は, 実際に得られた, または得られると予想される報酬の時間変化に対して応答することが示されたの

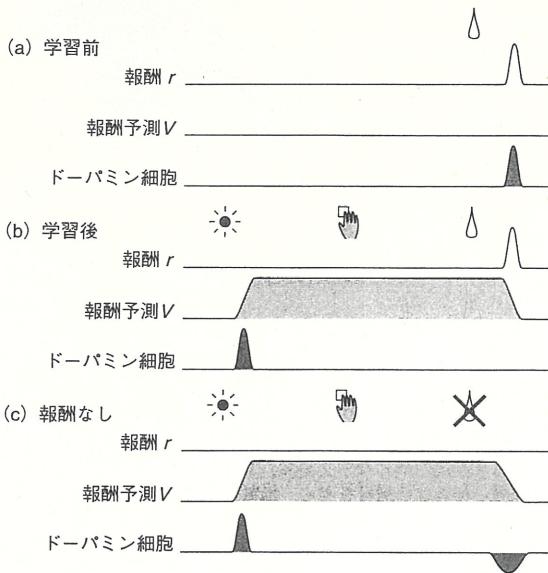


図4 大脳基底核の報酬予測的応答と、その強化学習モデルによる解釈。線条体では、将来得られる報酬の期待値 $V(t) = E[r(t) + r(t+1) + r(t+2) + \dots]$ が予測され、その予測と実際のずれを表わす“TD信号” $\delta(t) = r(t) + V(t+1) - V(t)$ が、ドーパミン細胞の活動として線条体での報酬予測の学習に用いられる。(a)学習前、あるいは手がかり刺激が与えられない場合、報酬予測は $V=0$ であり、TD信号は $\delta(t) = r(t)$ となり、ドーパミン細胞は与えられた報酬そのものに対して応答する。(b)学習が十分進んだ後は、評価 V は視覚刺激が与えられた時点で立ち上がり、報酬がもらえた後にゼロにもどる。TD信号は、刺激提示時点で評価の時間差分 $V(t+1) - V(t)$ によって正の値を取り、報酬が与えられた時点では、 V の時間差分の負の成分と、報酬 $r(t)$ の正の成分がうち消し合いゼロとなり、したがってドーパミン細胞は報酬ではなく、刺激提示の時点で応答する。(c)予測された報酬が実際には与えられなかった場合、評価の時間差分 $V(t+1) - V(t)$ の負の成分だけが残るため、TD信号、すなわちドーパミン細胞の活動は、減少する。

だ。

この実験データは、強化学習の理論モデル家からは非常に驚きをもって迎えられた。なぜならば、予測される報酬の時間変化というのは、強化学習のアルゴリズムの要となる重要な信号だからである。強化学習において最も大きな問題は、例えば、木によじ登って実を取ってきて、殻を割ったらようやくおいしいナッツが食べられるというように、複数のステップを経てからようやく報酬がもらえるような場合に、最後に得られた報酬の情報をもとに、いかにそこに至る一連の動作を学習するかという点にある。

これに対する理論的な解決策は、例えば木の実をみついたら20点、実を手にしたら50点、殻がうまく割れたら80点、といった具合に、報酬が得られる最終段階よりも前の状態に対して、内部的な評価を学習し、実際に報酬がまだ得られていなくても、評価が上昇するような行動はよい行動として強化する、というものである。そこで用いられるのが、(実際に得られた報酬)+(内部的評価の時間変化)という内部的な報酬として使われる量で、TD(temporal difference)信号と呼ばれている⁽⁹⁾。前述のSCHULTZらが記録したドーパミン細胞の活動は、まさしくこのTD信号そのものとみることができる(図4)。

この発見を契機に、大脳基底核の機能を強化学

習の枠組みで説明するモデルが提案されている(図5)⁽⁸⁾⁽¹⁰⁾⁽¹¹⁾⁽¹²⁾。それらによると、基底核の入力部である線条体の細胞は、大脳皮質からの入力をもとに現在の状態と、そのもとで可能な行動出力の評価を予測する。淡蒼球から視床、大脳皮質に至る経路では、最も高い評価値を与える行動出力が、競合の結果、選択される。黒質では、実際に得られた報酬と状態評価からTD信号が計算され、線条体へのドーパミン性の入力として送り返され、大脳皮質から線条体へのシナプスの可塑性を制御し、状態と出力の評価の学習に用いられる。

このような大脳基底核の強化学習モデルは、ドーパミン細胞の振舞いを説明するだけでなく、大脳基底核の回路に関する多くの実験的知見からも支持されている。まず、ドーパミン細胞の発火頻度がかつTD信号として強化学習に使われているとすると、線条体のシナプスの可塑性は、その入出力の相関によるだけでなく、ドーパミン入力の強さによって変化することが理論的に予想される。実際に予想どおり、黒質からのドーパミン線維は、大脳皮質から線条体の細胞へのシナプスの部分に結合し、ドーパミンの濃度によって、同じ入力条件でも、シナプス可塑性が長期抑圧(LTD)から長期増強(LTP)に変わることが示されている⁽¹³⁾。

線条体の尾状核には、特定の方向への瞬時的な眼球運動(サッケード)の準備中に活動する細胞が

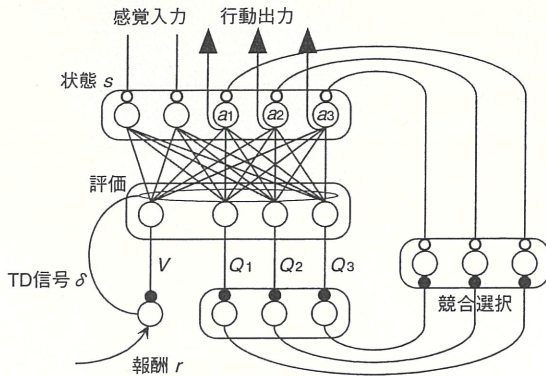


図5 大脳基底核の強化学習回路モデル(図3bとあわせてみてほしい)。身体、外界などの状態 s が大脳皮質に表現され、線条体の黒質に投射する部分では状態の評価 V が、淡蒼球、視床経由で大脳皮質に投射する部分では、行動出力の候補 a_1, a_2, \dots に対する評価 Q_1, Q_2, \dots が予測される。黒質では、実際に得られた報酬 r と、評価 V の増減をもとに、TD信号 $\delta(t) = r(t) + V(t+1) - V(t)$ が計算され、淡蒼球を経て、視床、大脳皮質への経路では、競合の結果、最もよい評価をもつ行動が選択され、大脳皮質から脊髄や他の脳領域への出力として送られる。TD信号 δ は、黒質からのドーパミン性の入力として、大脳皮質から線条体ニューロンへのシナプスに伝えられ、黒質投射部位では状態の評価 V の、淡蒼球投射部位では出力候補の評価 Q の学習に用いられる。

多くあることが知られている。最近、川越らは、上下左右の4方向のターゲットのうち、ある一方向への眼球運動でだけジュースがもらえるという条件で、尾状核の細胞の活動を調べた⁽¹⁴⁾。その結果、多くの細胞は、報酬がもらえる方向への運動の待機期間中に活動が増加することがわかった。報酬を与える方向を数十試行ごとに変化させると、細胞の方向選択性が報酬のもらえる方向に向かって大きく変化することが明らかになった。つまり、線条体の細胞は、単に特定の行動出力を固定的に表現しているのではなく、その行動の結果どれくらいの報酬が得られるのかという評価を与えるとみることができる。報酬条件に応じた細胞の特性の変化は、線条体の他の部位でも確認されている⁽¹⁵⁾。

これらの結果は、大脳基底核が、将来の報酬の予測にもとづき、現在の状態や、そこで取りうる行動の評価をおこない、それをもとにした目的指向的な行動の組立てにかかわっていることを強く示唆している。

大脳皮質の統計的教師なし学習

大脳皮質、とくに視覚野の細胞の刺激選択特性は非常に詳しく調べられ、例えば視覚一次野では特定の傾きをもった線分、MT野では特定の方向への動きなど、領域によってさまざまな特性をもつことが知られている。これらの性質は、各個体の視覚体験に依存して形成されることが示されており、それを説明するための教師なし学習のモデルが多数提案されてきている⁽¹⁶⁾⁽¹⁷⁾⁽¹⁸⁾。

教師なし学習の基本的な原理は、ある細胞がある入力によって発火すると、その入力に対するシナプス加重が増強されるという、いわゆるHEBB型のシナプス可塑性によって、つぎに同じような特徴をもつ入力信号がきたときにより強く応答するようになる、というものである。これに、各細胞の応答が強くなりすぎず、またそれぞれが異なる特徴に応答するように、相互抑制などの制御機構が組み合わせられる。教師なし学習には、クラスタリング、主成分分析、独立成分分析など、いくつかのバリエーションがあるが、いずれもHEBB型の可塑性と、双方向なダイナミクスによって実現可能である。

実際、大脳皮質には、各領域内、領域間で双方向の結合が非常に強く存在し、大脳皮質の細胞はHEBB型のシナプス可塑性をもつことが示されている。このことから、大脳皮質の重要な役割は、教師なしの統計的な学習の原理によって、感覚入力の中に潜む情報を抽出することだと考えられる。

ただし、われわれはみたもの聞いたものを何もかも等しく学習しているわけではなく、大脳皮質の学習が純粋な教師なし学習だと思うのには無理がある。大脳皮質でもとくに前頭連合野の場合は、その活動は単純な感覚情報ではなく、行動の意図や文脈に強く依存したものになっている⁽¹⁹⁾。

黒質のとなりの腹側被蓋野という部分のドーパミン細胞は、主に大脳前頭連合野に投射する。とくに前頭前野には、感覚刺激や行動プランを一時的に記憶にとどめておく“作業記憶”を担うニューロンがあることが知られているが、その記憶保

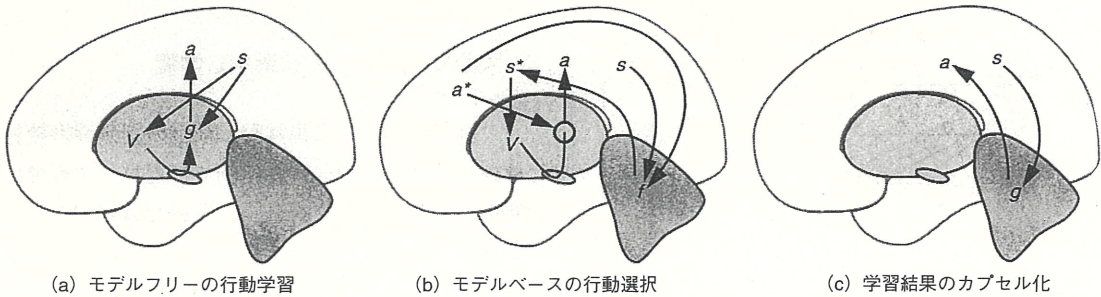


図6 行動学習系の基本的なアーキテクチャー。(a)モデルフリーの行動学習系では、状態 s に対して行動出力 a を取るという行動則 $a=g(s)$ がまず決められ、実際に行動してみた結果から状態の評価 $V(s)$ が大脳基底核で学習され、さらにそれをもとに行動則 g が更新される。(b)予測モデルにもとづく行動選択では、現在の状態 s のもとで、行動の候補 a^* を取った場合、身体、環境などの状態がどう変化するかを、すでに小脳が獲得している内部モデル $s^*=f(s, a^*)$ によって予測する。さらに予測された状態 s^* がどれくらいよいかの評価 $V(s^*)$ を大脳基底核が与え、候補である行動 a^* を実際に取るかどうかが決まる。このような行動選択をおこなないながら、内部モデル f 、評価 V が同時に更新される。(c)上記いずれの方式で学習された行動も、入出力関係 $a=g(s)$ を教師あり学習の原理によって小脳に記憶しておくことで、将来同じ条件のもとで、より素速く、脳のローカルな処理だけで実現することが可能になる。

持の強さが、ドーパミン入力のレベルに依存することが示されている⁽²⁰⁾。このことは、ドーパミンによる報酬予測の情報が、何を記憶にとどめておくべきかという表現の重みづけに使われていることを示唆している。

学習モジュールの統合

以上、小脳、大脳基底核、大脳皮質という、哺乳類で大きく発達した脳組織が、それぞれ教師あり学習、強化学習、教師なし学習という、学習のパラダイムの違いに応じて、特化した回路構造とシナプス可塑性のメカニズムをもつようになったという仮説を提示した。

昆虫やザリガニなどの無脊椎動物では、歩行や遊泳、飛翔など特定の行動に特化した神経細胞と回路構造が遺伝的に決定されている。それに対して脊椎動物、とくに哺乳類の神経回路では、特定の細胞や回路がどういった行動に関与するかは、生後の学習に強く依存する。おおざっぱに言えば、個別の行動に特化した神経回路を、短いライフサイクルでの自然淘汰によって最適化するのが無脊椎動物の神経系の適応戦略であり、個別の行動ではなく、特定の学習アルゴリズムを効率よく実現するために特化した神経回路を進化させ、生後の学習によってそれらを個別の行動のために割り当てていくのが哺乳類、ヒトへとつながる脊椎動物

の脳の適応戦略である。

このように考えると、脳の情報処理の理解のためには、ここは運動、ここは思考、ここは言語というように、単純に脳の場所ごとにラベルを貼っていくのではなく、ある行動の実現には、どのような学習要素が必要であるかをまず考え、そこから小脳、大脳基底核、大脳皮質の学習モジュールがどのように組み合わせられるべきかを推測し、それを脳の各部位の連絡関係と情報表現の知見とすり合わせ、それらの機能を考えるというアプローチが重要なことがわかる。

行動学習のアーキテクチャー

図6に、強化学習による基本的な行動学習のアーキテクチャーを示す。まず、最も基本的な区別として、“モデルフリー”と“モデルベース”の行動選択が考えられる(図6a, b)。

“モデルフリー”(モデルなし)の場合には、まずある適当な行動則を決め、それをもとに行動を試みる。その結果実際得られた報酬と、状態変化の経験をもとに、状態の評価が学習され、さらにそれをもとに行動則の学習がおこなわれる。これはいわば、あまり深く考え込まずとにかくいろいろやってみてから学んでいこうというやり方で、必ずしも賢くはないが非常にシンプルで確実な方法である。

このようなアーキテクチャは、大脳皮質に身体や環境の状態と、動作指令を表現する細胞が存在し、それらの評価値が大脳基底核で計算されるとすると、大脳皮質と大脳基底核をつなぐループ回路で実現が可能である(図6c)。実際、腕などの系列運動の学習では、大脳皮質の補足運動野とそこにつながる基底核の部位が強く関与していることが知られている。

いっぽう、“モデルベース”(モデルにもとづく)の行動選択(図6b)では、現在の状態での行動の候補を考え、それをおこなうと自分や環境の状態がどう変化するかを、すでに学習された内部モデルによって予測する。さらにその予測された状態がどれくらいよいものかを評価し、それが十分よければ想定した行動を実行に移し、よくなければ別の行動を考える。この方式は、予測モデルが現実とずれているとんでもない間違いを犯す可能性はあるが、予測モデルが通用する範囲内であれば、初めて体験する状況に関しても適切な行動が選べるなど、モデルフリーでおこなうよりも効率よい行動と学習が可能になる。

身体や環境、道具などの内部モデルが、前回示したように小脳に学習され、状態の評価が大脳基底核で学習されるとすると、モデルベースのアーキテクチャは、小脳と大脳基底核の連携動作によって可能になるはずである(図6b)。小脳と大脳基底核の間には直接の結合はないが、大脳皮質を接点とした小脳と基底核のループ回路によってこれは可能になる。実際、前頭前野や、運動前野の前方部は、短期記憶、運動の想像、暗算など、仮想的な状態の想像を伴う課題に強く関与している。

これらの部位は、小脳外側部との間にループ結合をもつだけでなく、大脳基底核にも投射を送っており、状態の予測とその評価という、モデルベースの行動アーキテクチャを実現するに十分な位置にある。これらの部位は、系列運動などの学習の初期に強く関与することが知られており、これをやったらどうなるかな、と考えながら行動を選択するという過程に関与していることが考えられる。

また、小脳の機能としてとくに重要と考えられ

るのが、行動則のカプセル化の機能である。とくにモデルベースのアーキテクチャでの行動選択は、小脳、大脳基底核、大脳皮質をつなぐグローバルな回路の直列的な処理が必要なため、計算時間の面でも脳の活動の面でも負荷が大きい。しかし途中の計算仮定は複雑でも、いっぺん望ましい行動則がわかってしまえば、その入出力関係を覚えておくことによって、つぎに同じ状況にきたときには迅速に判断を下すことができる。大脳皮質と小脳のループ回路は、身体や外界に関するモデルを提供するだけでなく、脳内で学習された入出力関係を、いわば脳の脳内モデルとして学習し、提供することが可能である⁽²¹⁾⁽²²⁾。

実際、非常によく習熟した運動や、暗算など、定型的な処理を高速におこなうという状況では小脳の活動が指摘されており、このようなメカニズムが、はじめは脳全体を使った意識的、直列的な処理から、無意識的、自動的、並列的な処理への移行過程と関連しているのかもしれない⁽²³⁾。

大脳皮質の状態表現

以上の議論では、大脳皮質には、小脳と大脳基底核をつなぐ単なるバッファの役目しか想定していなかったが、大脳皮質はおそらくそれ以上の意味をもつことが予想される。実際、予測モデルの学習、状態評価の学習の工学的研究では、状態や動作をどのような形で表現するかによって、学習の効率が大きく左右されることが知られている。

一般に感覚信号は、多くの冗長性をもつと同時に、不完全性をもつ。例えば、同じ物体も置かれる位置や光源によっていろいろな見え方をすると、空間上の同じ位置は、首や眼球の動きによって網膜上の異なる点に移る。それらをいちいちまったく別物だと思つと、予測や評価の学習には膨大な時間が必要だが、幸いにも、側頭葉や頭頂葉には、提示位置やサイズを変えてもある物体に不変に応答する細胞や、眼球位置によらず目標の位置を示すと考えられる細胞がみつけれられている。

最近では、1次視覚野の細胞でさえも、直接の視覚刺激だけではなく、トップダウン的に生成さ

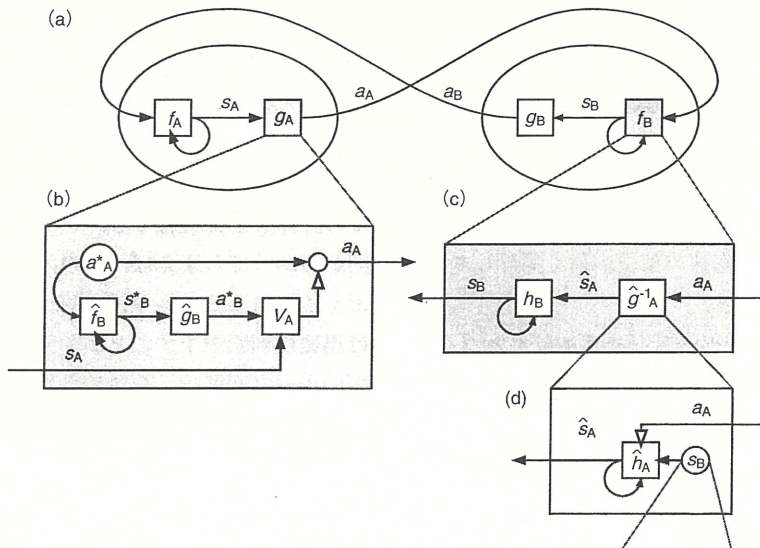


図7 コミュニケーションにおける内部モデルと評価機構の役割。(a)最も単純なコミュニケーションのモデル。Aさんの内部状態 s_A は、Bさんからの行動あるいはメッセージ a_B に応じて、 $s'_A = f(s_A, a_B)$ によって変化し、それをもとにメッセージ $a_A = g_A(s_A)$ が発せられる。(b)行動則 g_A の内部構造。メッセージを送るに際しては、あるメッセージ a^*_A を送った場合の相手の状態変化 s^*_B と、その結果の行動 a^*_B を、相手の状態変化と行動則の内部モデル \hat{f}_B と \hat{g}_B によって予測し、その結果が望ましいものかどうかの評価 $V_A(s_A, a^*_B)$ をもとに、メッセージ a^*_A を実際に発するか決定される。(c)状態変化 f_B の内部構造。BさんはAさんのメッセージ a_A に対して直接に応答するのではなく、Aさんの行動則の逆モデル \hat{g}^{-1}_A によってそのメッセージの背後にあるAさんの内部状態 s_A を推定し、それをもとに自分の状態を $s'_B = h_B(s_B, \hat{s}_A)$ によって変化させる。(d)逆モデル \hat{g}^{-1}_A の内部構造。一般に、メッセージ a_A に対する状態 s_A は一意には決まらないため、相手の状態推定 \hat{s}_A には、相手の状態変化 h_A の内部モデルを使った逐次推定 $s'_A = \hat{h}_A(\hat{s}_A, s_B)$ が必要になる。またその際に、相手は自分の真の状態を知っているとは限らないので、相手が想定しているであろう自分の状態 s_B を、AさんがもっているであろうBさんの内部モデルの内部モデルを使って推定する必要がある。

れた文脈情報に依存して活動することが示唆されており、冗長性を削減すると同時に文脈情報を補い、行動に必要な十分な状態表現を与えるということが、大脳皮質の非常に重要な役割であると考えられる。

また、情報を一時的に保持する短期記憶あるいは作業記憶のメカニズムは、上記のモデルベースのアーキテクチャにとくに必要なものであるが、さらに、短期記憶の容量が十分にあれば、つぎの状態の予測だけでなく、2ステップ先、3ステップ先の状態を予測し、それをもとに最適な行動を取ることが可能になる。チンパンジーでも、3ステップ程度までの先読みは可能であることが示唆されているが⁽²⁴⁾、深いステップの先読みが可能なのが、サル、チンパンジーとヒトの知能の根源的な差になっている可能性がある。これは、作業記憶を司るとされている前頭前野と、そこにつながる小脳外側部がヒトではとくに増大しているという事実と関連していると考えられる。

運動制御からコミュニケーションへ

ここまでは主に、単一の個体の行動を考え、ある動作をとると身体や環境の状態がどう変化し、どういう報酬が得られるかという枠組みで話をしてきたが、同じような発想は、複数の個体の間のコミュニケーションのメカニズムを考える上でも有効である。

まず最も単純な拡張は、言語などのコミュニケーションを、相手を制御対象とした制御の問題と考える立場である。例えば、単語は個々の動作指令、文は動作の系列であり、それに応じて相手の脳と身体の状態が変化する(図7a)。ここで、相手を自分の望む状態に動かすには、正しい動作を、正しい順序で与えなければならないというのは、制御の場合と同様である。

しかしこの発想は、相手が自分と同じ人間であるという事実を無視しているという点で、限界が

ある。実際、自閉症の患者には、相手の立場に立ってその人が何を考えているかを理解することが困難な傾向がみられることが指摘されている⁽²⁵⁾。

われわれが言葉を発するとき、失言癖のひどい人は別として、この言葉を発したら相手はどう思うか予測しながら、つまり相手の状態変化や応答の内部モデルを駆使して、単語とその系列を選んでいるのである(図7b)⁽²²⁾。もちろん、世の中にいる個々のすべての人に関する内部モデルをもつことは不可能なので、たいていの場合は、自分ならこう思う、という自分に関する内部モデルで代用することで対処する。

またわれわれが人の言葉に応答するとき、“危ない!”など単純かつ直接的な表現の場合を除いて、多くの場合は相手の言葉に表現された情景や気持ち(脳内状態)を想像する。つまり、コミュニケーション信号の受け手の側も、単純な固定的なダイナミクスをもつわけではなく、相手行動生成の逆モデルや、状態変化の順モデルをもち、その予測に対する評価をもとに応答しているのである(図7c, d)。

このような複雑なプロセス可能にするには、まず高度な内部モデルの学習機構、その予測を混乱せず保持する作業記憶の機構、さらにそこから適切な応答を選ぶための評価の機構がすべて必要であり、これには、言語野と呼ばれるような大脳皮質の一部だけでなく、小脳、大脳皮質、大脳基底核をつなぐグローバルな回路が関与していると考えられる。

さらに、人間の言語行動などを考えると、一連の系列で実現されるような行動や状態を、ひとつの単位行動や状態として定義する機能が重要な要素である。これには、小脳の教師あり学習による系列のカプセル化や、その活動を保持する大脳皮質の作業記憶のメカニズムが重要だと考えられる。今回は、状態予測と行動生成のモデルを並列的、階層的に組み合わせることによって、階層的な認

知行動機能がいかに実現されるかについて展開する。

文 献

- (1) K. DOYA: *Neural Networks*, 12, 961(1999)
- (2) 木村實・銅谷賢治: *科学*, 68, 970(1998)
- (3) F. A. MIDDLETON & P. L. STRICK: *Science*, 266, 458(1994)
- (4) F. A. MIDDLETON & P. L. STRICK: *Brain Research Reviews*, 31, 236(2000)
- (5) 銅谷賢次: “運動学習の神経計算機構—基底核, 小脳と大脳皮質”, 別冊・数理科学, 脳科学の前線, 141(1997)
- (6) J. C. HOUK & S. P. WISE: *Cerebral Cortex*, 2, 95(1995)
- (7) W. SCHULTZ et al.: *J. Neurosci.*, 13, 900(1993)
- (8) W. SCHULTZ et al.: *Nature*, 275, 1593(1997)
- (9) R. S. SUTTON & A. G. BARTO: *Reinforcement Learning*, MIT Press(1998)
- (10) J. C. HOUK et al.: in ‘Models of information processing in the basal ganglia’, J. C. HOUK et al. eds., MIT Press(1995) pp. 249~270
- (11) P. R. MONTAGUE et al.: *J. Neurosci.*, 16, 1936(1996)
- (12) H. NAKAHARA et al.: *Riken BSIS Technical Report*, 98-5(1998)
- (13) J. R. WICKENS et al.: *Neuroscience*, 70, 1(1996)
- (14) R. KAWAGOE et al.: *Nature Neurosci.*, 1, 411(1998)
- (15) M. SHIDARA et al.: *J. Neurosci.*, 18, 2613(1998)
- (16) C. VON DER MALSBERG: *Kybernetik*, 15, 85(1973)
- (17) S. AMARI & A. TAKEUCHI: *Biol. Cybern.*, 29, 127(1978)
- (18) B. A. OLSHAUSEN & D. J. FIELD: *Nature*, 381, 607(1996)
- (19) M. WATANABE: *Nature*, 382, 629(1996)
- (20) T. SAWAGUCHI & P. GOLDMAN-RAKIC: *J. Neurophysiol.*, 71, 515(1994)
- (21) M. ITO: *Trends in Neurosci.*, 16, 448(1993)
- (22) 川人光男: “小脳外側部の内部モデル”, 別冊・数理科学, 脳科学の前線, 194(1997)
- (23) O. HIKOSAKA et al.: *Trends in Neurosciences*, 22, 464(1999)
- (24) 松沢哲朗: *科学*, 69, 323(1999)
- (25) C. D. FRITH & U. FRITH: *Science*, 286, 1692(1999)