RIKEN BSI ITN Tech Report No.08-01-J.

題目：　大脳基底核; 報酬の予測と獲得のための強化学習(速習、概説)

著者名：中原　裕之（理化学研究所・脳科学総合研究センター・理論統合脳科学チーム）

１　大脳基底核の機能の理解のための計算論的視点

本稿では、大脳基底核の機能理解のための計算論的論点を整理して、紹介する。ある脳部位の機能を十分に理解しようと試みるとき、計算論的視点は重要である。計算論的視点とは、情報処理を理解するための視点と読みかえてもよい。その脳部位への入力と出力を考え、どのような情報が入力に存在し、どのような情報が出力に存在するか？その脳部位の、入出力の間で行う情報処理は何か、その処理はいかに実現されるのか？また、出力される情報は、脳全体の情報処理の中でどのような役割を果たすのか？そもそもなぜその脳部位がその情報処理を担当するのか、その回路特性はなにか？これらの問いかけは、ある脳部位の機能を十分に理解するための重要な問いである。例えば、「大脳基底核は、望ましくない運動を抑制します」としばしば言われるが、これは実は上記の問いかけに対する、簡単な答えかた、と言える。それは私たちの「言語的に記述した基底核の機能理解のモデル」である。計算論的研究は、この「モデル」を「“実際に動かすことができる”モデル」にすることを要求する。そのためには数理的モデルの構築が不可欠である。そして、その動くモデルができて始めて「理解のモデル」ができ、その脳部位の情報処理が理解でき、それにより機能の十分な理解を実現しようというアプローチが取られるのである。

計算論的視点から眺めたとき、大脳基底核回路の重要な特性が３つある。それは、１）基底核の基本的な回路特性、（２）大脳皮質―基底核間の並列回路構造、（３）学習信号としてのドーパミン神経細胞の存在、の 3 つである。紙面の限りがあるので、要点を概観することを中心に、簡潔にまとめておこう。より詳細については、本特集号のほかの章を参考にしてほしい。

（１）大脳基底核はいくつかの神経核から構成される。その大きな特徴は、神経核から他の核への投射神経細胞が、主として、抑制性神経細胞で構成される点にある。特に、基底核の主たる投射を受ける神経核を入力核、基底核から他の脳回路へ投射を送る神経核を出力核と呼ぶことにすると、この入力核と出力核はともに他へは、抑制性の投射をする（図1)。入力核の自発発火活動は低い一方、出力核の神経細胞のそれは高い。従って、自発レベルでは、基底核の出力は、投射先の神経細胞活動を抑制するのが主たる働きと考えられる。入力核から出力核への経路は、入力核から出力核への直接の投射が存在し、これは直接経路と呼ばれる。（本稿では別の経路、例えば間接経路などの経路、についての説明は省く。）従って、入力核の神経細胞が強い発火活動を行うときには、この直接経路を介し、出力核の活動を抑制する。出力核の投射先の神経活動を考えると、この入力核の出力核への抑制は、脱抑制として機能し、出力核の投射先の神経活動を上昇させることになる。これ

が抑制―脱抑制の機能を持つといわれる所以である（図１）。基底核から主要な出力経路は、‘下流’の脳幹運動中枢へむかう経路と、視床を介して大脳皮質に投射をもどす二つの経路に大別される（図 2)。ここで、この両方向への投射は、大脳基底核が運動出力の調節、特に、自律的あるいは本能的とも言える運動出力と、多様な情報処理を行った後で決定できる運動出力の両者を調節しつつ、最終的な行動選択や運動出力を決定づけるのに適した部位だと考えられる。

（２）大脳基底核は、大脳皮質の広い領野から投射を受けている。皮質からの主要な入力経路は、線条体と側坐核、つまり基底核回路の入力神経核への投射である。（この入力核は、他にも視床などからも投射を受ける。一方、大脳皮質からは視床下核などの基底核回路の途中の神経核へも投射がある。)。基底核から出力経路のうち、大脳皮質に戻る経路は、基底核の入力核へ入力を送っていた広い大脳皮質に投射を戻している（図３)。従って、大脳皮質―大脳基底核―視床―大脳皮質というループ構造をなしている。ここで、特徴的なのは、このループ構造が、幾つかの並列なループに分類できることにある）（図３)。

（３）大脳基底核回路に存在するドーパミン神経細胞が、「快」の刺激あるいは「報酬や報酬刺激」に対して反応することは良く知られていた。しばしば引用される知見は、脳内刺激に関する知見と、嗜癖(addiction)に関する知見である。近年、実は、その神経活動の詳細が分かってくるにつれ、むしろ、ドーパミン神経細胞活動は、報酬予測の誤差を表している、さらには、それが学習信号として使われるということを示唆する知見が積み上げられつつある。

本稿では、第２・３節で、計算論的考察から得られる強化学習という視点から、このドーパミン神経細胞に関する知見をまとめる。それに基づき、第４節で、大脳皮質―大脳基底核の並列回路との関係について議論したうえで、まとめることとしたい。

２　ドーパミン神経細胞の報酬誤差をあらわす活動

ドーパミン神経細胞が豊富に存在する領域は、黒質緻密部と腹側被蓋野である（図３)。大脳基底核の入力核である線条体と側坐核は、共にドーパミン神経細胞の投射を豊富に受ける。線条体は、さらに尾状核、被殻に分かれる。（注：線条体が、尾状核、被殻、そして側坐核に分かれるという整理のしかたも実は存在する。）尾状核と被殻は主に黒質緻密部からドーパミン細胞の投射を受ける。その一方、側坐核、そして側坐核の出力先の腹側淡蒼球は、主に腹側被蓋野からドーパミン細胞の投射を受ける。これらの入力核は、全般的に基底核の抑制－脱抑制の機能から、意思決定・行動選択・運動制御に全般的には関わると考えられる。相対的には、尾状核はより意思決定に、被殻はより運動出力に関連が深く、一方、側坐核は、扁桃体や視床下部との深井関係から、より情動や感情の関わる行動選択に関連が深いと考えられる。ドーパミン細胞がひとたび強く活動すると、これらの脳部位に一斉にその活動が伝えられ、その影響は大脳皮質－基底核の並列回路を通じて脳全体に

及ぼされることになる。このように基底核全体、ひいては大脳皮質の活動に影響を及すことが可能で、また、「快」や「報酬刺激」に関連して活動するとされるドーパミン神経細胞の機能を理解するのは、大脳基底核の機能解明の大切な基礎になる。

近年の研究では、このドーパミン神経細胞の活動が報酬予測誤差に対応する活動を表すことがシュルツ（W. Schultz）らの研究を契機に明らかにされつつある。典型例としては、古典的条件付課題と呼ばれる実験課題、つまり、条件刺激として「光」があって、その一定時間後に報酬が与えられるサルの実験がある（図４A)。ドーパミン神経細胞は、最初は（学習前には）報酬そのものに反応する。この光―報酬の試行を何度も繰り返すと、ドーパミン神経細胞はやがて（学習後には）報酬そのものには反応しなくなる。その一方で、報酬を予期させる条件刺激としての光に反応をする。さらに、この学習後に、報酬を省くとその活動は抑制を示す。このように、ドーパミン神経細胞の活動は、期待する報酬からのずれを表すことが分かってきた。このような実験的発見の積み重ねに導かれて、ドーパミン神経細胞活動の機能を理解するのに、現在「報酬予測誤差仮説」が有力な仮説となっている。ここで、簡単には、報酬予測誤差とは、

報酬予測誤差　＝　実際の報酬　－　期待した報酬　　（1）

と考えればよい。次節では、この仮説の理論的バックボーンとなっている強化学習について概観しよう。

３　報酬予測誤差を利用した学習――強化学習

報酬予測誤差は、何の役に立つのか？大いに役立つのである。報酬という用語は、そもそも「対価」と「強化因子」という二つの意味を持っている。対価は、まさしくその場で得られるものを、強化因子はそれによって後々に同じ場面で同様の行動を引き起こすようになることを、各々意味する。報酬予測誤差は、強化因子に適している。誤差が正ならば、予測よりも得られたものが大きいということだから、それに対する強化は続けられるべきだろうし、誤差がゼロになれば、予測どおりということだからそれ以上強化はいらないだろう。負の誤差は予測よりも負の誤差が少ないことを意味するので、むしろその選択を弱めるように働くだろう。このような考えに基づいて古くから定式化されているのが、条件付け学習においてよく知られているレスコラーワグナーの学習ルール(Rescorla-Wagner learning rule)である。ここで紹介する強化学習は、アルゴリズム・数理の両面でより精錬された学習方式で、ドーパミン神経細胞活動とより広く対応することが分かっている。

強化学習は、機械学習とか人工知能で計算機の上に実装できるかなり強力な学習アルゴリズムである。応用例としては、例えば、ロボットの制御、オセロやバックギャモンのゲームなどがある。バックギャモンでは、人間の世界チャンピオンと同じくらい強くなることができる。実は、強化学習の学習アルゴリズムにはいくつか種類があり、ドーパミン神経細胞の研究では、主として temporal difference learning と呼ばれるアルゴリズムが用いられる（以下、TD 学習と呼ぶ)。TD 学習での学習信号は、TD 誤差（TD error)と呼ばれる。

ドーパミン神経細胞活動がこの TD 誤差でよく説明されることが、報酬予測誤差仮説の理論的裏づけとなっている。

時刻 t の TD 誤差を $\delta(t)$ で表すと、それは

$$\delta(t) = r(t) + \gamma V(t+1) - V(t) \qquad (2)$$

と数式で表現される。右辺で、$r(t)$ は時刻 t の報酬、$V(t)$ は時刻 t の期待報酬を表す。γ は未来の報酬の割引率を表す。先ほどの例で報酬時を考えてみよう（図４B）。学習前は、$V(t) = V(t+1) = 0$ なので、$\delta(t) = r(t)$ という TD 誤差つまりドーパミン神経細胞の反応があると考えられる。一方、学習後は、上の式を $\delta(t) = r(t) - (V(t) - \gamma V(t+1))$ と書き直してみると分かりやすいが、報酬 $r(t)$ と、期待報酬の二つの時刻の差分（時間差分）である $(V(t) - \gamma V(t+1))$ が釣合うことで、$\delta(t) = 0$ となり、ドーパミン神経細胞の反応がなくなる。一方、このような時に報酬を与えないと、$\delta(t) = 0 - (V(t) - \gamma V(t+1)) < 0$ という抑制された反応が現れてくる。報酬を予期させる条件刺激提示時でのドーパミン神経細胞の反応も同様に説明することができる。学習前では、式（２）の右辺の各項は全てゼロなので、TD 誤差はゼロになる。一方、学習後では、刺激提示により始めて期待報酬がゼロから変化する、つまり $V(t) = 0$、$V(t+1) \neq 0$ だから（そして、もちろん、この時刻では $r(t) = 0$）、結果的に TD 誤差はゼロにはならず、ドーパミン神経細胞のような正の反応が現れる。

この TD 誤差は、学習信号として利用される。例えば、期待報酬を表す関数（価値関数）$V(t)$ が、

$$V(t) = \sum w_i s_i$$

として表されている場合を考えよう。右辺の s_i は、各時刻の状態を表していると思えばよく、w_i はそれに対する重みづけと思えばよい。このとき、TD 誤差 $\delta(t)$ を利用して、この重みづけは、

$$w_i \rightarrow w_i = w_i + s_i(t)\delta(t)$$

と変化（つまり学習）していく。この学習を繰り返すことで、各時刻での期待報酬を正確に予測できるようになる。

TD 学習でのこの価値関数はなかなか巧妙にできている。$\delta(t) = 0$ となっているときを考えながら、式（２）をぐっと睨んでもらうと分かるのだが、この価値関数は、学習が十分行われた後では、

$$V(t) = r(t) + \gamma r(t+1) + \gamma^2 r(t+2) + \gamma^3 r(t+3) + \cdots$$

と表せるのである。一般的な状況では、何度も行動選択をしたあとでやっと報酬が得られることも多い。そんな場合に目前の報酬だけ考えて行動選択をすることは決して賢い方法ではない。このような時間遅れがある報酬に対処するには、目先の報酬と遠い将来の報酬の両方を考慮に入れる必要がある。上の価値関数は、同じ量の報酬でも将来へと先延ばしになればなるほど、割り引かれるようになっている。このように、異なる時刻に与えられる報酬を全て考慮しながら、各時刻 t での期待報酬を表しているのが上の価値関数なのである。

価値関数を学習することで、各時点において、将来の報酬を含めた期待報酬を正しく予測できるようになる。この報酬の正しい予測が、動物の報酬獲得のための行動選択に役立ちそうなことは直観的に理解できる。この価値関数を利用しつつ、得られる報酬を最大化するために行動選択そのものを学習すればいいのである。この行動選択関数の学習については紙面の都合で詳しく述べる余裕がないが、要諦は、TD 誤差を学習信号として利用しつつ、実現されることにある。

4　大脳皮質―大脳基底核回路の機能

前節で、ドーパミン神経細胞活動の報酬予測誤差仮説の理論的裏づけである TD 学習を簡潔に説明した。この仮説に基づけば、ドーパミン神経細胞活動が TD 学習の学習信号である TD 誤差に対応する。この視点から、大脳基底核の機能の理解がどのように広がるか考えていこう。論点整理のために、基底核での TD 学習、皮質―基底核の機能分化、皮質―基底核並列回路の各ループ、そして最後にループ間の協調と相互作用と、各々の観点から眺めていこう。本来これらの観点は、互いに関連しあうので、この順に分離して議論できるものではないが、ポイントを最初に概観するやり方としては有効だと思う。

まず第一に、TD 学習の枠組から眺めてみると、学習信号（TD 誤差）がドーパミン神経細胞ならば、それでは価値関数や行動選択関数が脳のどこに存在するのかが重要になる（強化学習の枠組としては、他にも行動価値関数などがあるが、ここでは最も簡潔なこの二つの関数だけを考えよう）。ドーパミン神経細胞の主たる投射先でなければ、TD 誤差を学習信号としては使用しづらい。その点では、線条体・側坐核、つまり基底核の入力核が、価値関数や行動選択関数の主たる候補となる。逆に、ドーパミン神経細胞が存在する黒質緻密部と腹側被蓋野への主たる投射を送る領野を考えると、ここでも入力核が候補に挙がる。このように線条体・側坐核は、これらの関数が脳の機能として実現されるのに大きな役割を果たしていると考えられる。

一方で、入力核だけで全て行われると必ずしもいいきれない。例えば、TD 誤差の計算には、価値関数を正の値として与えることも必要である（式２の右辺第二項）。入力核からの黒質緻密部と腹側被蓋野への直接投射は抑制性であり、興奮性にするには途中の経路を介する必要がある。図 1 に一つの可能性を示しておいた。それは黒質網様部から緻密部への投射である。これは入力核の神経活動からみると、二つの抑制を介することで脱抑制の機構として働く。したがって興奮性の投射に近い機能を実現できる。もうひとつは、興奮性の直接投射を送る領野として、脚橋被蓋核は有力な候補である。最新の実験研究でそのような報告がなされていて今後の研究の進展が待たれる。また、入力核だけで機能が完成すると考える必要もない。入力核へ投射を送っている大脳皮質領野も価値関数や行動選択関数の形成に大いに役立っていると考えられる。

第二に、大脳皮質―大脳基底核回路における皮質と基底核の機能分化としては二つの点

が注目される。一つは、同じ実験課題で調べると、皮質領野と基底核の両方で報酬関連応答がある。しかし、概ね基底核のほうが、そこに入力を送る皮質領野よりも、報酬による神経細胞活動の修飾が強い。これは、皮質－基底核全体の情報処理としては、基底核のほうがより報酬関連の情報処理の役割を担っていることを示唆する。やや単純化しすぎた見方かとは思うが、「認知」と「情動」と峻別したとき、皮質はより「認知」に近く、基底核はより「情動」に近いと考えられる。もう一つは、解剖的特徴や、あるいは大脳皮質の領野間での投射を考えると、大脳皮質のほうが、基底核よりも、より複雑な情報処理に適した構造を有する。実際、同じ実験課題で調べると、概ね大脳皮質のほうが、その投射先の基底核よりも、神経細胞活動は多様性に富んでおり、また、報酬以外の実験条件に応答する活動も多い。これは、たとえば前節で強化学習のアルゴリズムとして TD 学習を紹介したが、この TD 学習は比較的単純なアルゴリズムとなっている。それに対し、より複雑だが、より計算能力の高い強化学習のアルゴリズムも数理的には存在し、そのような情報処理が大脳皮質で実現されている可能性を示唆している。

第三として、皮質－基底核並列回路の各ループにおける機能分化と、その各々の特性をいかしたループ間の協調と相互作用が考えられる。まず機能分化としては、各ループに属する大脳皮質領野と基底核領野は、各々異なる神経活動の特性を持っている。たとえば、主に視覚刺激に反応する領野は比較的同じループに属し、一方で、眼球運動に主に反応する領野が属するループや、腕の運動出力に反応する領野が属するループも存在する。このように各ループで表現されている情報にはそれぞれ特徴がある。これらの特性を生かした協調と相互作用を考えるときに着目すべきになのは、各々の情報表現に違いがある一方で、学習信号としては、広く投射するドーパミン神経細胞から、ほぼ同様の学習信号を受けていることにある。この特徴に着目して、私たちは多重表現仮説を皮質－基底核回路の基本的な機能として 90 年代の終わりに提唱した。この仮説は以後の研究でも概ね支持を受けていると考えている。その基本骨子は、情報表現の性質を反映して、同じ運動系列を学習していても各ループで獲得される学習形質が異なる、それがひいてはループ間の協調に役立つというものである。たとえば、同じ運動系列でも視覚座標系の表現のほうが学習が早くできる一方、関節座標系で表現していると学習が遅くなる。このとき、運動系列を早く学ぶループ（「視覚ループ」、これは図 4 では背側前頭ループと表記されている）と遅く学ぶループ（「運動ループ」）があり、共に強化学習に基づいて学習する。ここで、学習早いほうは、その時々で新しい系列を学習し、制御に貢献する。一方、遅いほうはなかなか貢献できないが、実は、学習が遅いぶん、ゆっくりとではあるが確実に学んでいく。ひとたび十分な学習をするとその記憶はなかなか消えない。しかもひとたび学習されれば、関節座標系で表現されているため、その実行はすばやくできる。このように本当に重要な運動系列は、ゆっくりとした学習の後では、主として運動ループを介して制御が可能になる。愚鈍でもその特性の生かし方はあるのだ。

以上駆け足で、皮質―基底核回路の計算論的特性を、ポイントに分けて見てきた。

上述した各々のポイントが全体としてどのようにして優れた情報処理を実現しているのかについては、まだ分かっていない点も多く今後の計算論的研究が待たれている。これらの研究が報酬誤差仮説、そしてその理論的背景にある強化学習の研究の進展と共に進むことは間違いないであろう。私たちもそれらの研究に貢献したいと日夜努めている（最新の研究については、研究室のＨＰ www.itn.brain.riken.jp をぜひご覧ください。）今後の大脳基底核の機能解明の研究が大きく発展することを期待している。

参考文献

Alexander GE, DeLong MR, and Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience* 9: 357-381, 1986.

Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, and Doya K. Parallel neural networks for learning sequential procedures. *Trends Neurosci* 22: 464-471, 1999.

Hikosaka O, Nakamura K, and Nakahara H. Basal ganglia orient eyes to reward. *J Neurophysiol* 95: 567-584, 2006.

Hikosaka O, Takikawa Y, and Kawagoe R. Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol Rev* 80: 953-978, 2000.

Kobayashi Y, Inoue Y, Yamamoto M, Isa T, and Aizawa H. Contribution of pedunculopontine tegmental nucleus neurons to performance of visually guided saccade tasks in monkeys. *J Neurophysiol* 88: 715-731, 2002.

Nakahara H, Doya K, and Hikosaka O. Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences - a computational approach. *J Cogn Neurosci* 13: 626-647, 2001.

Nakahara H, Itoh H, Kawagoe R, Takikawa Y, and Hikosaka O. Dopamine neurons can represent context-dependent prediction error. *Neuron* 41: 269-280, 2004.

Schultz W, Dayan P, and Montague PR. A neural substrate of prediction and reward. *Science* 275: 1593-1599, 1997.

図のレジェンド

図 1：大脳基底核の基本的な回路構成について、眼球運動系を例に示す。**(A)** 各領野から他領野への投射を示すために、主たる投射細胞をマルで示し、その投射先へのシナプス結合を長方形で示している。白丸は興奮性神経細胞、黒丸は抑制性神経細胞を示す、投射末端での＋またはーは、各々興奮性と抑制性の結合があることを示す。大脳基底核の入力核である尾状核は、大脳皮質から興奮性の結合を受け、抑制性の投射を出力核である黒室網様部に送る。この出力核は、眼球運動制御で主要な役割を担う上丘に抑制性の投射を送る。黒質緻密部にはドーパミン神経細胞が豊富に存在し、それを斜線のマルで示している。このドーパミン神経細胞は尾状核の神経細胞に投射している。大脳皮質から大脳基底核への投射のシナプス結合の可塑性を修飾することを＊で示している。このシナプス可塑性の修飾機能が、本文で説明している強化学習を実現すると考えられている。**(B)** 典型的な神経細胞活動のシークエンスを示す図。二重に存在する抑制性結合の活動をもとにした抑制―脱抑制の連関が、上丘で起きていることを示している。(Hikosaka, Nakamura, Nakahara 2006 J. Neurophysiology の図をもとに改変。)

図 2; 大脳基底核からの運動関連の主たる出力経路。(Hikosaka , Takikawa, Kawagoe 2000 Physiological Review の図を改変)。

図３；大脳皮質―大脳基底核回路の並列回路を示すダイアグラム。この回路は、皮質―基底核―視床―皮質の順に巡るループ回路が基本的な構成である。このループは、主としてさらに幾つかの並列なループに分けられる。図中の①から⑤は各々並列ループを模式的に示している。(Alexander, DeLong, Strick 1986 Ann. Rev Neurosci の情報をもとに作成。Nakahara & Hikosaka in preparation の図を改変。)

図 4 ; 刺激－報酬課題（古典的条件付課題）におけるドーパミン神経細胞活動 **(A)**とそれに対応する TD 学習における TD 誤差 **(B)**。（B）の下の 2 行の図は、ドーパミン神経細胞活動とＴＤ誤差が一致することを、報酬がもらえる場合ともらえない場合の両方について示している。

図 1

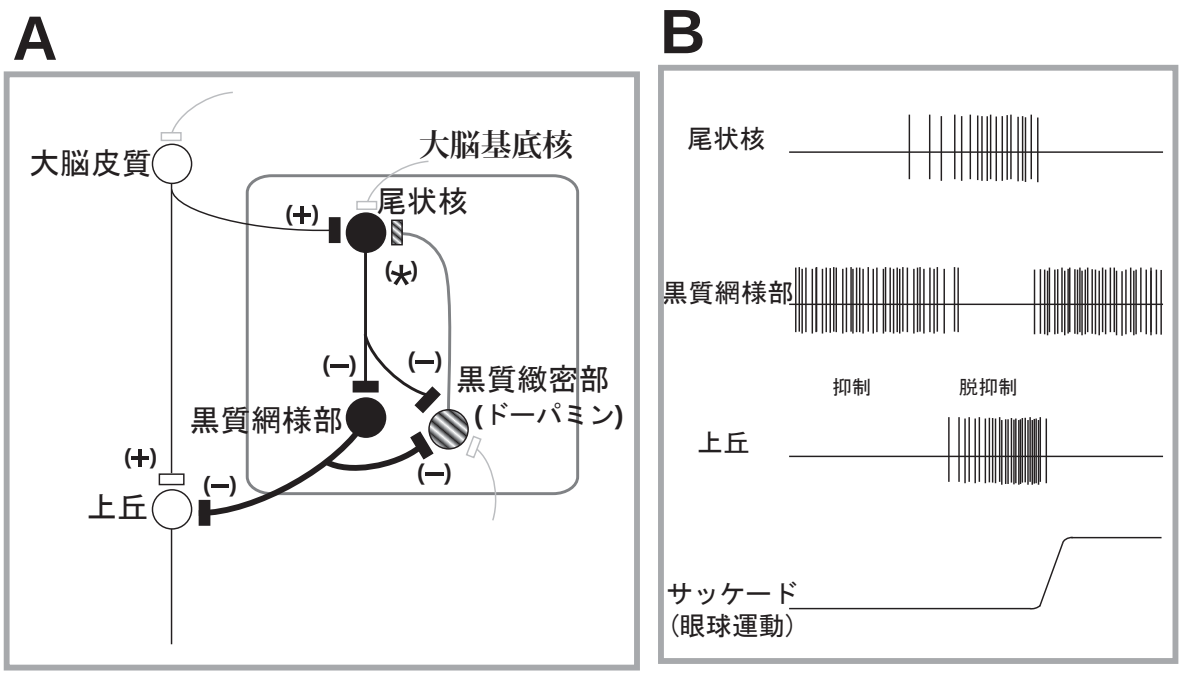

A
大脳皮質
大脳基底核
尾状核
(+)
(*)
(−)
(−)
黒質緻密部
(ドーパミン)
黒質網様部
(+)
(−)
(−)
上丘
B
尾状核
黒質網様部
抑制
脱抑制
上丘
サッケード
(眼球運動)

図 2

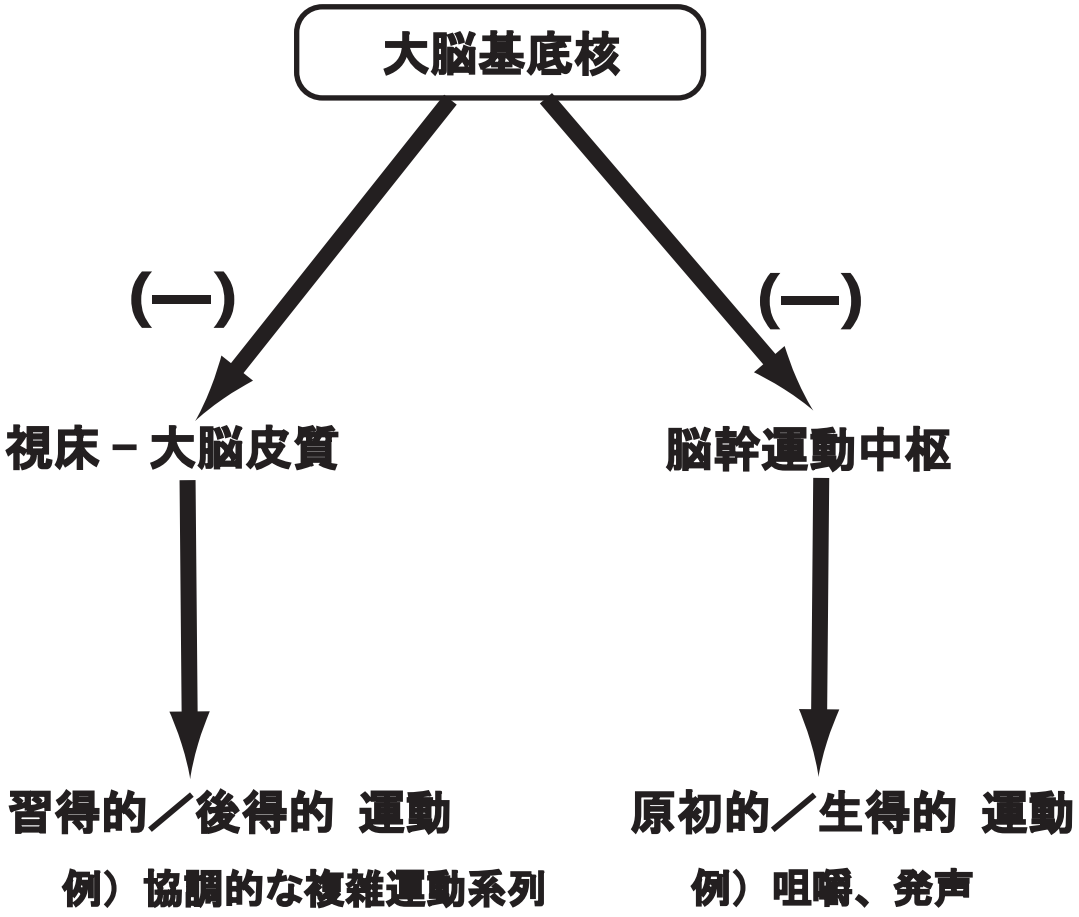
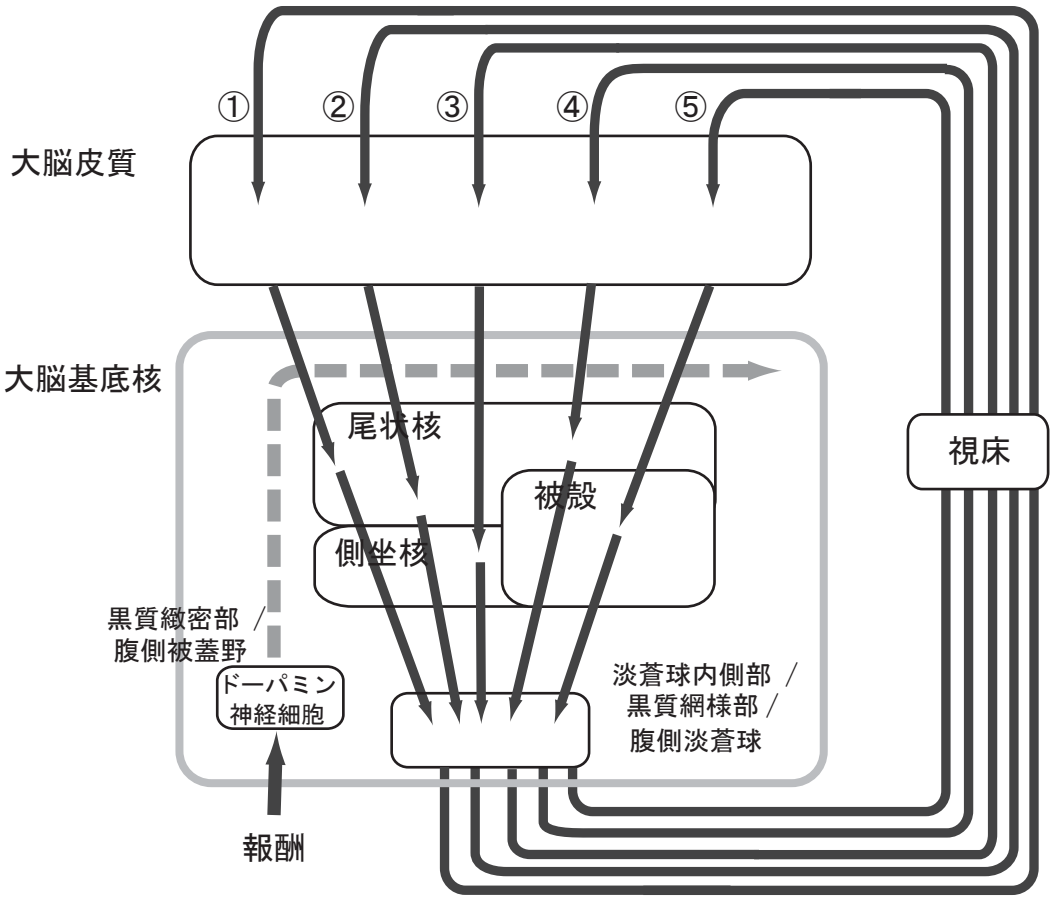


図 3



① 背側前頭ループ

② 外側眼窩前頭ループ

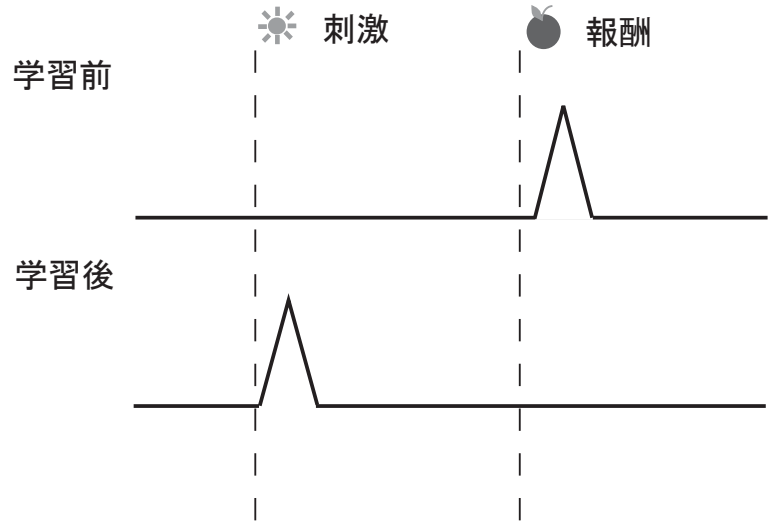
③ 前帯状皮質ループ

④ 眼球運動ループ

⑤ 運動ループ

図 4

A　ドーパミン神経細胞（DA）活動



B　TD 学習；DA 活動と TD 誤差