変化する脳：進化/発達/学習/修飾

銅谷賢治
doya@atr.co.jp
ATR 人間情報科学研究所／科学技術振興事業団 CREST

1　神経情報科学サマースクールのねらい

近年の分子生物学，機能イメージング手法，計算機の発展を土台として神経科学が大きく進歩している．しかしその結果得られる大量の詳細な実験データが何を意味するのか，あるいはどういう情報処理に関連するのかを明らかにするためには，データを理論モデルに基づき解析し，データと理論によるシミュレーション結果と比較したりすることが求められている．このような背景のもと，数理モデルを構築し解析の出来る実験家，実データを扱い実験を企画出来る理論家の育成を目的に，このサマースクールは企画された．

本スクールでは，各テーマに関する概論だけではなく，その分野の最先端の現状と課題を提供することを目指す．また，ただ講義を聞くだけでなく，グループ討論や計算機実習，仮想研究提案，講義録作成を通しての主体的な参加を求めている．これらの体験を通して，異分野の研究者間の交流と，学際的なネットワークが形成されることを期待している．

2　各講義の内容

今回のテーマは，一言で言えば脳の適応機構であるが，適応と言ってもさまざまレベルのものがある．数億年の単位で推移し主に遺伝子によって制御される進化，数年から数日単位での個体の神経回路の経時的変化である発達，数日から数分単位のシナプス結合の変化とみなされる学習，数日から数秒単位で修飾物質によってもたらされる脳内物質による修飾が考えられる．今回のスクールではこれらの共通メカニズムと差異，それらの間の相互依存関係について考えることとしたい．ここではまずこの全体的なテーマと各講義の内容を概観する．

進化

まず岡ノ谷先生には生態学の概念を一般的に導入した後，先生が行われている鳥の歌の学習における行動と遺伝的要因の対応について説明して頂く．

山元先生には遺伝子と行動の関わりをショウジョウバエの性行動を対象として解説して頂く．

今年はヒトゲノムがほぼ解読された記念すべき年であるが，今後はこの膨大なデータからいかに情報を読み取るかが重要なテーマとなる．浅井先生には統計的な学習手法を用いてこのデータから情報を抽出する手法について紹介して頂く．

昆虫の神経回路の進化を調べておられる下澤先生は昆虫が熱雑音のレベルまでのセンシングを行うメカニズムについて説明して頂く．

発達

まず桝先生に分子生物学的立場から誘導物質によって神経回路が形成されるメカニズムを紹介して頂く．

田中先生には発達の研究が最も進んでいる視覚野に関して自己組織化がどこまで視覚体験に依存するかに関するモデルと光記録による実験について紹介して頂く．

富田先生には細胞1個をまるごと計算機上でシミュレーションするE-Cellプロジェクトについて解説して頂く．

学習

学習とシナプス可塑性について，海馬とならんで深く研究が行われている小脳に焦点を絞り，理論と実験の共同研究の成功例を見ることとしたい．まず川人先生に眼球運動を学習を中心に小脳の学習に関して理論的側面を紹介して頂く．

平野先生には小脳の長期増強と長期抑圧の分子メカニズムについて解説して頂く．

また，これまで小脳の学習の分子機構に関しては物質間のブロック図は描かれてきたがその定量的検証はあまり行われてこなかった．黒田先生にはkinetic simulationを用いたシグナル伝達経路のシミュレーション研究についてご紹介頂く．

修飾

日常生活の経験からも分かるように脳の活動は注意や意欲という要因に大きく左右される．これは同じ神経回路が神経修飾物質の働きかけに応じて機能を変化させることで起こる可能性がある．今回は主にドーパミン，セロトニン，ノルアドレナリン，アセチルコリンの役割について考えたい．

曽良先生にはノックアウトマウスを用いたドーパミン，セロトニンの相互作用と薬物依存の分子メカニズムについてご紹介頂く．

臨床医でいらっしゃる瀬川先生にはモノアミン系の物質が担う機能特に子供の脳の発達に対する影響について解説頂く．

最後に石井先生には修飾物質の機能と考え得る学習におけるランダムさの自動調節について主に計算論的な立場から計算機シミュレーションを交えて解説して頂く．

Figure 1: 起立運動学習ロボット.

3　小脳/基底核/大脳皮質の学習

以下では今回のスクールに関連する2つの話題について述べる．ひとつは小脳，大脳基底核，大脳皮質における学習について，もうひとつは脳内修飾物質の働きに関する強化学習の立場からの仮説についてである．

小脳と基底核は従来運動制御のみに関与する器官であると考えられてきた．ところが近年の臨床，機能イメージング研究の進歩からこれらが認知機構にも関連することが次第に明らかになってきた．MiddletonとStrickによるヘルペスウィルス逆行トレーサーによる実験で，小脳の出力核である歯状核，大脳基底核の出力核である淡蒼球が高次機能を司る前頭連合野の46野に強い出力を送っていることが分かった[1]．これは小脳，大脳基底核が運動制御だけでなく高次の認知機能にも深く関連していることを示している．

そこでこれらの部位がそれぞれ何をしているかが問題となるのであるが，ここでは学習アルゴリズムの観点から考える．

3.1　大脳基底核と強化学習

強化学習とは得られる累計報酬量を最大化する様に行動戦略を学習する計算パラダイムのことである[2]．近年，大脳基底核の主な機能は強化学習に強く関与しているという仮説が提唱されている．

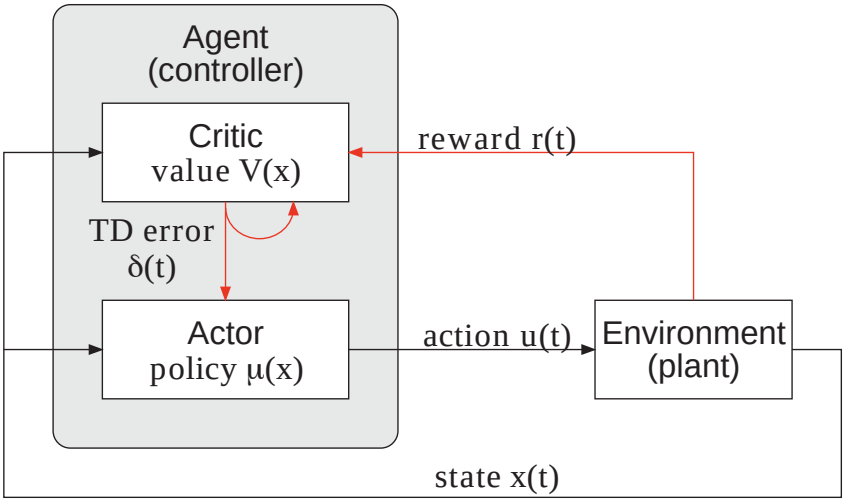


Figure 2: Actor-Critic学習方式.

図 1は，強化学習により起立運動を学習ロボットである [3]．身長70cm, 体重5kgでジャイロセンサを搭載している．報酬としては頭の高さを用いることで，このロボットは試行錯誤で次第に起き上がることを学習する．ここで重要なのは，もしロボットが現在の頭の高さのみを最大化しようとすると上体だけ持ち上げるようなつまらない行動だけを学習してしまう．つまりうまく立ち上がる為には，長期的な報酬を最大化する行動を取ることが必要となるのである．そのためには，どのように学習を進めれば良いだろうか？

理論的には報酬の予測という概念を用いればうまく学習を進められることが分かっている．図 2はActor-Crtic と呼ばれる学習方式である．これは，現在の環境とその時点で得られる報酬から将来に渡る報酬予測を行うCriticと，現在の環境とCriticの予測から行動$u(t)$を決定するActorから構成される．そして，Criticの報酬予測誤差であるδ(t)が，Critic，Actor両者の学習の鍵として用いられる．具体的には，Criticは出来るだけ正確な報酬予測を行うようにδ(t)を小さくするように学習を進め，Actorは得られる報酬を増大させるためδ(t)を大きくする行動を選択するように学習を進める．

3.2 価値関数とTD誤差

では，報酬の長期予測や報酬予測誤差を具体的に定義してみよう．強化学習のエージェントは状態x_1で行動u_1を取り，報酬r_1を得る．この時状態はx_2に推移するものとする．以後同様に，時刻tに状態x_tで行動u_tを取り，報酬r_tを得て，状態はx_{t+1}に推移する

状態x_tから見た長期の報酬予測を

$$V(x_t) = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots]$$

と定義し，状態価値関数と呼ぶ．価値関数は状態x_tからスタートして平均的にどうの程度の報酬が得れるかを示している．γは減衰率で0と1の間の値を取る．この値は将来の報酬をどの程度重要視するかを決定する．

同じ式を$V(x_{t+1})$について書きくだし，その差を取ることにより，2つの状態の価値関数の間には

$$V(x_t) = E[r_t + \gamma V(x_{t+1})]$$

という関係が成り立つべきことがわかる．ここからのずれ

$$\delta_t = r_t + \gamma V(x_{t+1}) - V(x_t)$$

は，長期的報酬予測の誤差を表し，Criticの出力の学習に使われる．$\delta(t)$はTD誤差と呼ばれる．これはまた，当初の予想に比べてどれだけ多く報酬が得られたか，または長期的に見て有利な状態にたどり着いたかを示す信号であり，その前に取った行動に対する強化信号，つまりActorの学習信号としも用いられる．

3.3 ドーパミン細胞の報酬予測応答

大脳基底核が強化学習に重要な役割を果たしているのではないかという仮説のもととなったのは，Schultz et al.の一連の実験である．サルに，ランプがついた後，正しくレバーを押したら報酬（ジュース）を与えるという課題を行わせながら，中脳ドーパミン細胞の活動を記録したものである（文献[4], Fig. 1参照）．

未学習時は，ドーパミン細胞はサルの口にジュースが与えられられた時に反応している．学習が進み，ランプがついたらレバーを引いてジュースをもらえることがほぼ確実にできるようになってくると，報酬に対する反応は減ってきて，代わりにランプがついたことに対する反応が大きくなってくる．さらにその後で報酬をカットすると，ドーパミン細胞の活動が抑えられる．これらの結果は，ドーパミン細胞は，報酬そのものではなくて，報酬の予測誤差δに対して応答しているのではないかということを示唆している．

学習前は，報酬予測Vはどの状態でもゼロだとすると，TD誤差$delta = r$で報酬そのものに等しい．学習が進むと，ランプがついた時点で報酬の予測値Vが増加するので，報酬r

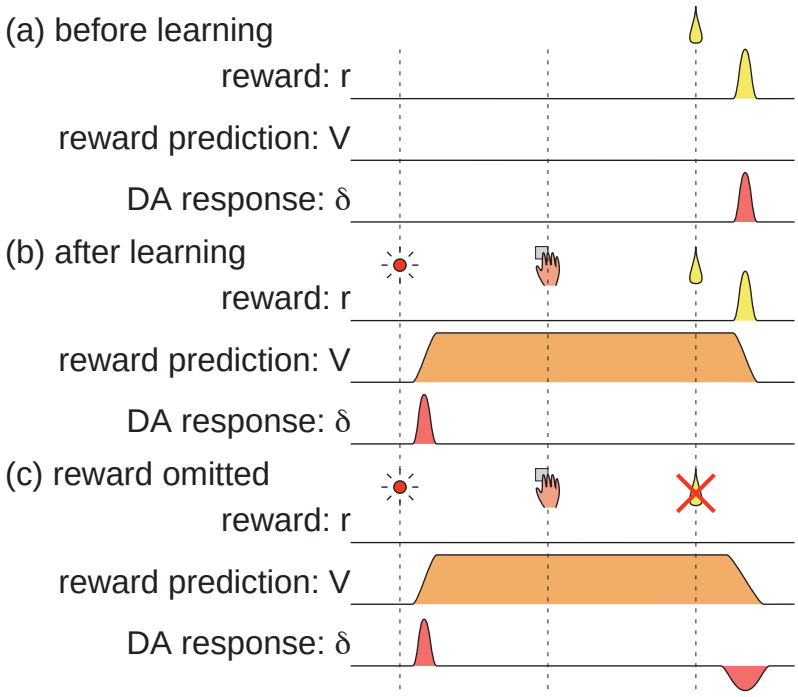


Figure 3: ドーパミン細胞の報酬予測活動の TD 学習モデル.

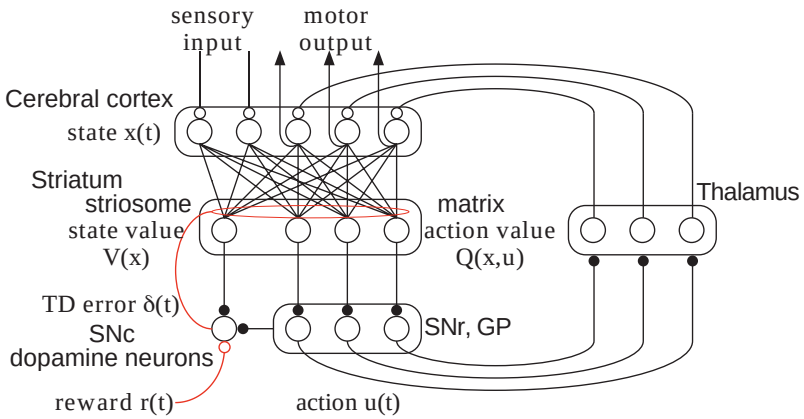


Figure 4: 大脳基底核の強化学習回路モデル.

はゼロでも，$\delta_t = V(x_{t+1}) - V(x_t) delta$ の微分信号によTD誤差は正の値を取る．レバーを押し報酬がもらえると，実際にもらえた報酬に対する r の正の値と，今後は当面報酬は来なくなるための V 微分の負の値が相殺されて，$\delta_t = 0$ となる．もし報酬 r が来ないと，V 微分の負の成分だけが残り δ はマイナスになる．

3.4 大脳基底核の強化学習モデル

黒質ドーパミン細胞の主要な投射先は，大脳基底核の入力部である線条体である．Schultzらの発見を機に，線条体やそこへ信号を送る大脳皮質の細胞の報酬に関連した応答の記録がさかんに行われている．例えばKawagoeらによる眼球運動中線条体の活動記録は，線条体ニューロンの発火は単にこれから取る行動だけでなく，その結果どれだけ報酬が得られるかによって大きく変化することを示している [5].

図 4は，これらの事実をもとに提案された大脳基底核の機能モデルである．線条体のうち黒質ドーパミン細胞に投射する部分 (striosome) は，大脳皮質に表現された状態 x をともに価値関数 $V(x)$ を計算する，つまりCriticとして働く．その出力と，視床下部などからの報酬そのものの情報をもとに，黒質ドーパミン細胞でTD誤差 δ が計算される．

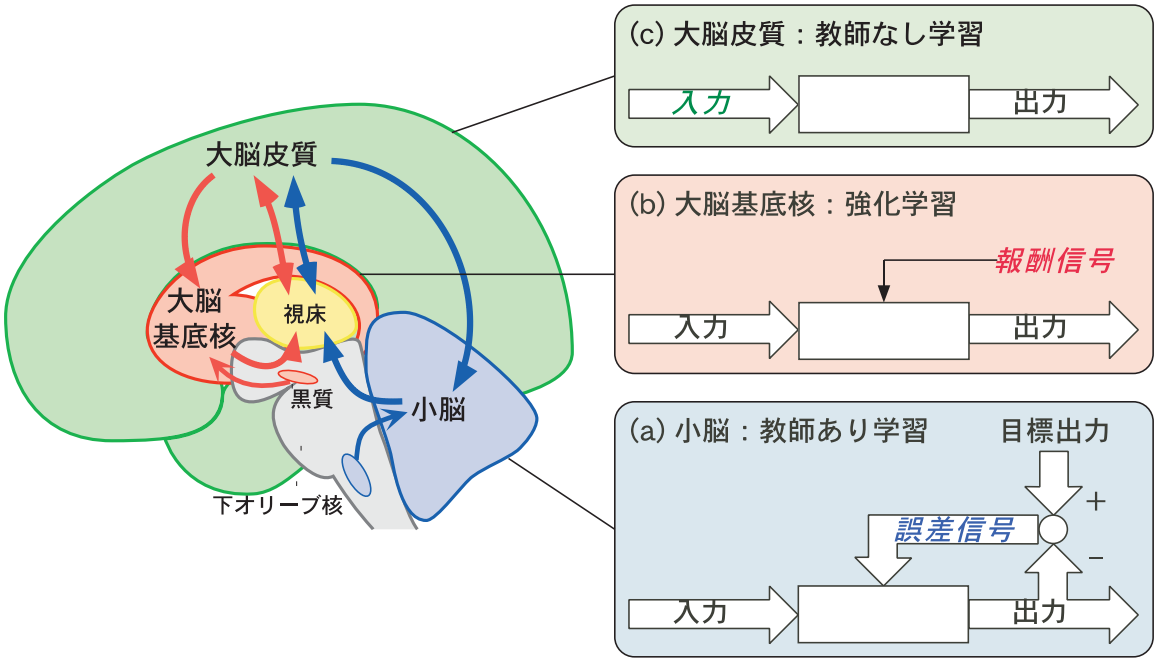


Figure 5: 小脳，大脳基底核，大脳皮質の学習アルゴリム．

線条体のうち黒質網様部や淡蒼球を経て視床や脳幹の運動関連部位に信号送る部分 (matrix) は，その場所ごとに異なる行動 a を取った後に予測される報酬を表し，その出力の競合により行動出力が選ばれる．つまり，Acotr として働く．

ドーパミン細胞の線条体へのフィードバックは，TD 誤差をもとにした，線条体ニューロンの学習に使われる．

3.5　学習パラダイムごとの専門化

初めに述べたように，小脳や大脳基底核の機能は，単に運動制御というだけではとらえきれない．大脳基底核の回路は，これまで見てきたように，強化学習を実行するのに適した構造をしている．

では小脳はというと，その回路構造とシナプス可塑性をもとに，登上線維入力を誤差信号とした「教師付き学習」を行うという理論が古くから提案されている．また大脳皮質の細胞の様々な応答特性は，感覚信号の統計的性質に依存した「教師なし学習」の枠組みから説明されている．

教師付き学習，強化学習，教師なし学習というのは，学習の理論でいう最も基本的な３つの学習の枠組みである．多くの実験データとモデルは，小脳，大脳基底核，大脳皮質が，これらの異なる学習の枠組みにそれぞれ特化したシステムだということを示唆している（図 5）[6].

昆虫など無脊椎動物の神経系は，ほ乳類などに比べたら非常に少ない数の神経細胞で，実

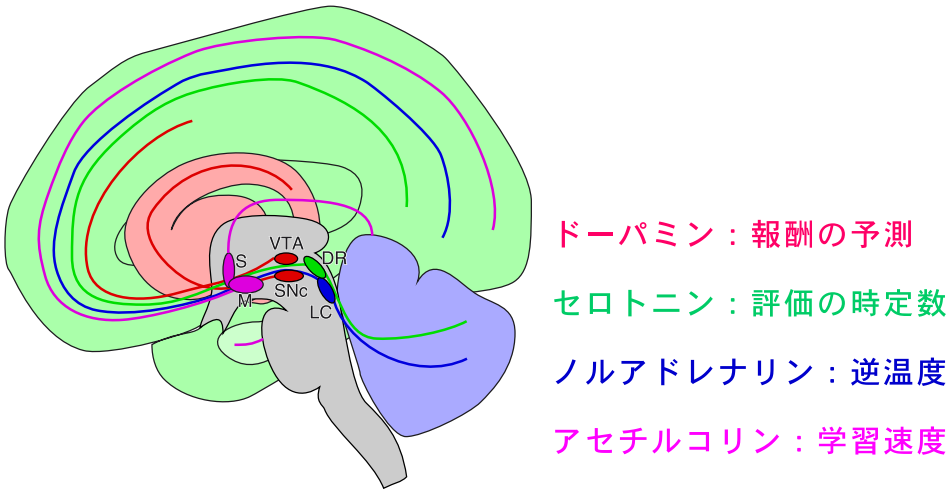


Figure 6: 代表的な神経修飾物質系．黒質緻密部（SNc）と腹側被蓋野（VTA）から投射するドーパミン系．背側逢線核（DR）からのセロトニン系．青斑核（LC）からのノルアドレナリン系．中隔核（S）とマイネルト核（M）からのアセチルコリン系．

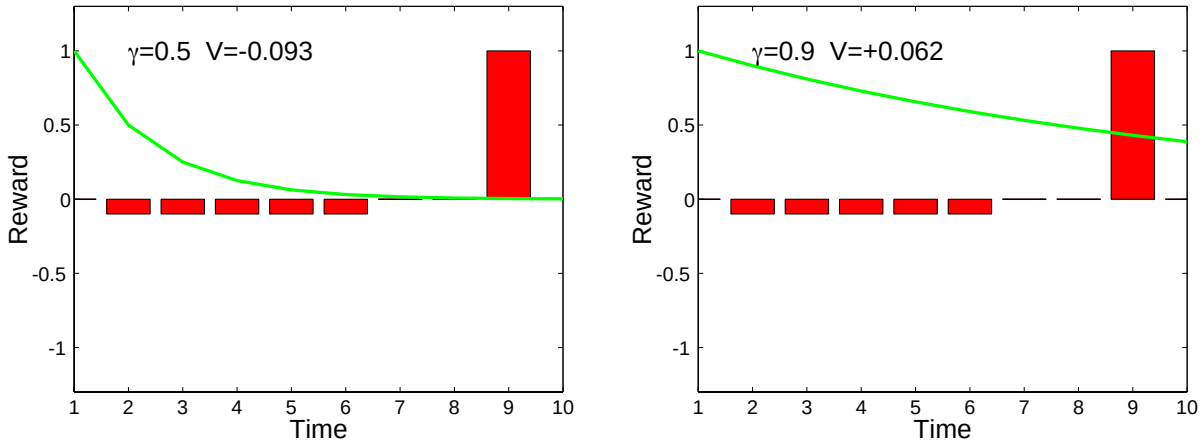
に巧みに行動制御を行っている．これは，生存に必要な行動レパートリーに特化した神経回路を，短いライフサイクルで多数の子孫を作ることで遺伝的に進化させたことによる．それに対して脊椎動物，特にほ乳類の脳は，特定の行動ではなく，学習のアルゴリズムに特化した神経回路を持ち，生後の学習により適応を図るという，対照的な適応戦略と取っていると言える．

4　神経修飾物質系の計算モデル

実際に強化学習のシミュレーションやロボット実験をしてみると，学習がうまく進むためには，学習アルゴリズムの中に出てくる係数の設定が非常に重要であることがわかる．例えば，行動探索のランダムさん，報酬予測の時間スケール，記憶の更新のスピードなどである．これらは，システムの動作を決める多数のパラメタが，学習により変化するしかたを決めるパラメタであることから，メタパラメタと呼ばれる．

先に見せた学習ロボットなどでは，学習のメタパラメタは実験者が経験と勘を頼りに決めてあげている．動物や人間の学習では，外から誰かがそれを調整してくれているのでないとすると，脳自身が，メタパラメタを調整する，メタ学習の機能を持っているはずである．

図 6は，脳幹から小脳，大脳基底核や大脳皮質に投射する，代表的な神経修飾物質系を示したものであるが，これらが，脳におけるメタ学習の機能を司っていることが考えられる [7].



(a) (b)

Figure 7: 割引率γの設定による行動評価の違い.

4.1 報酬予測の減衰係数γ

ドーパミンが強化学習で最も重要な学習信号であるTD誤差δを表しているのではないかということはすでに述べた.

価値関数の定義に出てくる，将来の報酬評価の減衰係数γは，学習される行動を左右する重要なメタパラメタである．将来大きな報酬を得るためには，つらい労働を日々重ねなければいけないというのはよくある状況である．図 7(a) のようにγが小さいと，将来の報酬への重みがほぼゼロになり，目先の負の報酬だけが評価に入り，こんな苦労はしない方が良いということになる．図 7(b) のようにγが小さいと，将来の報酬が十分評価されるため，当面負の報酬を払っても先の報酬を目指した行動が学習される．

セロトニン系の活動低下は，うつ病や自殺，衝動的な行動との相関が見られる．このことは，報酬予測の時間スケールの制御にセロトニンが関わっていることを示唆している．

4.2 温度と学習速度

強化学習では，様々な行動を確率的に試してみる必要がある．しかしそのランダムさを決めるメタパラメタは，熱力学とのアナロジーで「温度」と呼ばれるが，これは学習の度合いに応じて調節する必要がある．ノルアドレナリン系は，その活動が高いほど行動のランダムさが減る，つまり温度の逆に相当するメタパラメタを制御していることが示唆されている．

また，学習によるパラメタ更新の速度係数も重要なメタパラメタである．アセチルコリンは，海馬のシナプス可塑性や回路の動作に影響を与えることから，以前の記憶を保持／再生するか，新しい情報を記銘するかという，学習のゲート信号の役割を担っていることが示唆されている．

4.3　神経修飾物質系のメタ学習仮説

以上をまとめると，哺乳類の代表的な神経修飾物質系は，強化学習における以下のような変数やメタパラメタを表現しているのではないかと考えられる．

ドーパミン：報酬の予測（デルタ）
セロトニン：評価の時定数（ガンマ）
ノルアドレナリン：逆温度（ベータ）
アセチルコリン：学習速度（アルファ）

これらの変数やメタパラメタの設定には相互依存関係があり，また環境条件や行動経験に応じて変化すべきものである．それらに関する理論的な予想から，脳の物質系の複雑な相互作用を理解することが可能になるかもしれない．

5　まとめ

脳の適応戦略として，無脊椎動物は行動に応じた回路の最適化，脊椎動物は学習アルゴリズムに応じた最適化をしていると考えることができる．

神経修飾物質系は，行動モードの選択や行動学習をメタレベルで制御しているのではないかと考えることができる．

質疑応答

Q:学習時のガンマの値の与え方についていい方法はあるのか？

A:まだそれについては決定的な答えは見出されていないが，一般的に学習の初期段階においては，あまり先のことを予測できないだろうから，ガンマの値を抑えて，それからだんだん値を上げていくという方法をとると高いパフォーマンスが得られている．具体的にどのような式やどのような値を代入すればいいのかなどについては分かっていない．

Q:ロボットがある行動を獲得した時に見られるシナプス結合は，再現性のあるものなのか？

A:最終的に獲得される入出力の写像のようなものは似ているが，シナプス結合のパターンという意味では毎回かなり違ったものになると思う．

Q:報酬の与え方も重要だと思うのだが...

A:脳の中でも報酬を獲得するためにどれだけ犠牲になってもいいかという制御はおそらくされていて，今の強化学習の枠組みだとプラスの報酬とマイナスの報酬は単純に足し合わせてしまっている．報酬の評価と危険の評価をそれぞれ別に学習しておいて，状況に応じて安全戦略をとるとかある時には一か八かの選択をとるとかいうことはかなり重要な問題だと思う．

参考文献

[1] Middleton F.A., Strick P.L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. Science,266,458-461.

[2] Sutton R.S., Barto A.G. (1998). Reinforcement Learning: An Introduction. MIT Press.

[3] 森本淳, 銅谷賢治 (2001). 階層型強化学習を用いた３リンク２関節ロボットによる起立運動の獲得. 日本ロボット学会誌,19,574-579.

[4] Schultz W., Dayan P., Montague P.R. (1997). A neural substrate of prediction and reward. Science,275,1593-1599.

[5] Kawagoe R., Takikiwa Y., Hikosaka O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. Nature Neuroscience,1,411-416.

[6] Doya K. (1999). What are the computations of the cerebellum, the basal ganglia, and the cerebral cortex. Neural Networks,12,961-974.

[7] 銅谷賢治 (2000). 行動学習系のメタパラメタ制御と神経修飾物質. 数理科学,38,19-24.