

大脳基底核の計算モデル

講 師：中原 裕之

平成 12 年 8 月 10 日

1 大脳基底核と強化学習

強化学習の基礎と彦坂先生の基底核の実験について既に今日の講義で聞いたと思うので、まず最初に、強化学習と大脳基底核との関係についてモデルの立場から手短かに述べたいと思います。その後で、今日は我々が行ってきた二つの研究を紹介したいと思います。一つは、先程の彦坂先生の実験に関連して、報酬信号に基づく自己組織化という観点から基底核の細胞の反応に関するモデル研究で、もう一つは、基底核の機能を、これから説明する Actor-Critic モデルの立場から考えたときの大脳皮質-基底核回路がどのように逐次運動制御・学習を実現しているかという研究です¹。

では、まず、TD 学習と、Actor-Critic モデルの基底核との関係を話しましょう。Actor-Critic モデルでは、TD 誤差という強化信号を使い、Critic モジュールと、Actor モジュールがそれぞれ学習を行います。Critic は各状態の評価関数の学習を目的とし、TD 誤差を 0 にするよう学習を行う。

¹ この前半の研究は、甘利俊一（理研）・彦坂興秀（順天堂大）との共同研究で、後半は、銅谷賢治（ATR）・彦坂興秀（順天堂大）との共同研究で、それぞれの研究の出典は、

- Self-organization in the basal ganglia with modulation of reinforcement signals. Nakahara H, Amari S, Hikosaka O., submitted for a journal publication.
- Parallel Cortico-Basal Ganglia Mechanisms for Acquisition and Execution of Visuo-Motor Sequences - A Computational Approach. Nakahara H, Doya K, Hikosaka O., Journal of Cognitive Neuroscience (in press).

です。なお、本文中にいくつか文献が引用してはありますが、他にも引用文献はあって、現状の引用は不完全であることを予めお断りしておく。いずれ整理した上で、引用文献・図表・定理の証明などについては、つけ加えることにする。この講義録は、森本淳（奈良先端大）・石田文彦（東北大）によって起こされました。お二人に感謝します。

その一方、Actor は誤差信号をプラスの方向に最大化するよう各状態でのアクションを選ぶように学習を進めます。この二つの学習で結果的には、一番よい系列を選ぶように学習が進んでいきます。

Schultz らのドーパミン細胞に対する実験結果をごく簡単に説明すると (Schultz et al.,1993)、彼らの実験では、reward の前に reward が来ることを知らせる刺激 (conditional stimulus;CS) を与えます。条件付けが不十分だと、reward が与えられた直後に中脳ドーパミン細胞の phasic な反応がある。その一方、十分に条件付けが行われると、ドーパミン細胞は CS に phasic に反応し、reward への反応は消える。さらに、CS の前にもう一つ別の CS を与えると、その刺激にドーパミン細胞は反応し、元々の CS への反応は消える。

このドーパミン細胞の反応は先に述べた TD 誤差によく似ています。というのも、学習前には各状態の評価関数は獲得されていないので、TD 誤差は reward のところに現れます。ところが、評価関数の学習後は、TD 誤差は刺激のところに現れる。この関係が大脳基底核と強化学習、とくに TD 学習との対応付けの議論のきっかけとなった (REF)。もちろん以前から、パーキンソン病、ハンチントン病など、逐次系列の行動障害の事例や、ドーパミン細胞は元々 reward 刺激に対する反応が強いなどの状況証拠があったことも見逃せない。

さて、Actor-Critic モデルと大脳基底核との対応 (Houk et al.,1995) はというと、Actor にあたる線条体内の細胞は現在の状態に応じた運動指令の選択を行い、淡蒼球と視床を経て大脳の運動関連領野へ出力を送る。一方、Critic にあたる線条体内の細胞は状態の評価を計算し、ドーパミン細胞へ投射する。ドーパミン細胞はこの評価信号ともともとの報酬信号 (reward) との差をとって、TD 誤差を作り出し、線条体に拡散的に投射している。これを利用して Actor と Critic は学習を行うという仮説がたてられた。この仮説は大変有望であると考えられているが、その一方で、例えば、実際に Actor, Critic の場所は基底核のどの部分に対応するのかなど、まだ不完全・未確認の点も多い。

- Q: この対応がうまくいくには、ドーパミン細胞からの拡散的な線条体への投射は強さの情報をもっていないといけないのか?
- 中原: 理論的には確かにその通りだ。しかし、今のところドーパミン細胞の phasic な反応と予測信号の予測誤差についての反応との量的な対応を示すデータはない。厳密な意味ではないが、Schultz らの上記の例ではドーパミン細胞の反応は時間軸上をさかのぼると、

多少は小さくなっている。

- Q: 彦坂先生の例では, reward の方向に対するドーパミン細胞の反応についての特性は示されているので, 加えて大きさの情報があれば, 対応付くのでは? そのようなデータはないのか?
- 中原: おおまかには対応付くと言えるが, きちんとした実験データは, 今のところない。これから説明する彦坂先生の caudate の実験では, all-directions-rewarded condition (ADR) と one-direction-rewarded condition(1DR) で1ブロックの reward の総量は一緒になるようにしている。この時, ADR の各方向の reward の量を1DR のときの1/4 になっている。ニューロンの反応はADR の方が1DR に比べ弱くなっているという結果はあるけれど、しかし、これが厳密に reward 情報の量的な表現があるかどうかはまだ不明である。強化学習で単純に考えると reward 信号は Markov 過程だが、実際には Markov 的ではなく、他の方向の reward 信号も関わってくると考えられる。

では、一つめの研究の紹介を始めましょう。

2 報酬信号の調節下での caudate ニューロンの自己組織化

Kawagoe らは、非対称に reward を与える memory-guided saccade task を用いて, reward の modulate による caudate (尾状核: 大脳基底核の神経核のうちで、大脳皮質からの入力を受ける線状体の前部にあたる神経核のこと) のニューロンの反応の変化を調べた (Kawagoe et al.,1998)。memory-guided saccade task では、注視点を中心に提示され、次にキュー刺激がスポットライトとして、4方向のうち1方向に与えられる。そして、注視点が消えた後、サルはキュー方向に saccade を行う。60 試行を1ブロックとし、1ブロック中、全方向に reward を与える all-directions-rewarded condition(ADR) と、4方向のうち1方向にだけ reward を与える one-direction-rewarded condition(1DR) の2つ異なる reward 条件で、サルにタスクを行わせた。1DR では、異なる実験ブロックでは異なる方向に reward を与えた。なお、reward を与えない方向でも、サルは正しく saccade を行わなければならない。行わないと、同じトライアルをくり返

される。1DR は4ブロック行い、各ブロックでは reward 方向が異なっていた。

この実験結果の全体については、現在も解析が進行中だが、Kawagoe et al.,1998 で報告された結果を簡単にまとめると、まず第一に、各々の1 DR ブロックの reward 方向の違いに応じて、caudate ニューロンの各キュー方向の入力への反応特性は、ブロックごとに変化する。次に、その reward 方向により modulate されるニューロンの反応の変化の仕方はおおまかに3タイプに分かれることが分かった。その3タイプとは、

a).flexible type

最も多く見られ、ニューロンの示す選択的な方向は各1DR ブロックの reward 方向へすばやく変化する。

b).conservative type

各ブロックを通して、選択的な方向を保ち、その方向が reward 方向のとき一番強く反応する。

c).reverse type

他の2タイプより少なく、各1DR ブロックの reward 方向に対しての反応が一番弱い。flexible type と逆の反応を示す。

である²。

さて、少し話を戻して、大脳基底核の一般的な次の3つの特徴に注目しよう。

- 1). 大脳基底核は大脳皮質のほぼ全体から入力を受け、上丘、前頭葉にその出力を戻すが、大脳皮質→大脳基底核経路のニューロン数を比較すると、だいたい、大脳皮質10000ニューロンに対し、大脳基底核は100ニューロン程度である。
- 2). 大脳基底核内の線条体には、抑制系のニューロン (co-lateral や interneuron) が豊富である。

² ちょっと寄り道すると以前、このニューロンの各タイプを Actor-Critic と対応付けようと試みた。例えば、Critic が各条件で reward に対する反応、予測をだしているならば、Critic は flexible type のニューロンであろう。各方向に対応する Actor が競合して役割をはたすなら、Actor は空間選択性が高く、reward modulation がある。基底核の視床への direct な経路と indirect な経路があることも考えると、conservative type が direct な経路の、reverse type が indirect な経路のそれぞれの Actor であろう (Nakahara et al., 1998; Trappenberg et al 1998)。

3). 線条体は黒質緻密部のドーパミン細胞から豊富な強化信号を受けている。

この3つの特徴をふまえた上で、我々は大脳基底核の線条体で強化信号に modulate された自己組織化が行われるというモデルを考えた。というのも、この自己組織化により、1) の convergence があっても、3) の報酬信号を利用して、皮質からの入力に対し効率のよい情報表現を作ることができる。しかも我々のモデルでは、2) の抑制性の細胞をうまく利用できるなど、これら基底核の特徴を上手く利用できるからである。

実は、我々のモデルは、甘利らが70~80年代に提案したモデル (Amari,1977, Amari and Takeuchi,1978, Amari,1983) をベースに考えた。甘利らは70年代に抑制系の可塑性を利用した視覚野の自己組織化モデルを提案した (Amari,1977, Amari and Takeuchi,1978, Amari,1983)。

$$y = f(u(t))$$

$$u(t) = \mathbf{w}(t) \cdot \mathbf{x}(t) - w_0(t)x_0(t)$$

y, u はそれぞれニューロンの出力、内部状態、 \mathbf{w}, w_0 は興奮性、抑制性の荷重値、 \mathbf{x}, x_0 はそれぞれ興奮性、抑制性の入力、 f は出力関数である。興奮性、抑制性の荷重値の学習則は

$$\tau \dot{\mathbf{w}}(t) = -\mathbf{w}(t) + cf(u(t))\mathbf{x}(t)$$

$$\tau \dot{w}_0(t) = -w_0(t) + c'f(u(t))x_0(t)$$

ここで、 c, c' はそれぞれ興奮性、抑制性の荷重値の学習に対する学習係数である。第一項は、発散を抑える項で、第二項は Hebb 学習の匂うである。このモデルでは、方位選択性を示す、入力情報に対する効率のよい表現が得られる、などの特徴があることが明らかにされた。

線条体 (ニューロン) の出力を y 、大脳皮質からの入力を \mathbf{x} 、皮質と線条体間の荷重値を \mathbf{w} と書くことにする。ドーパミン細胞は reward を受けるとドーパミンシグナル α を計算し、 α が \mathbf{w} を modulate する形のモデルを考える。ニューロンモデルは

$$y(t) = f(u(t))$$

$$u(t) = \mathbf{w}(t) \cdot \mathbf{x}(t)(1 + \alpha(t)) - w_0(t)x_0(t)$$

で、 u はニューロンの内部状態、 f は出力関数、 w_0, x_0 はそれぞれ抑制性の荷重値、入力を表す。簡単のため、ここでは抑制性入力 x_0 を定数とす

る。学習則は

$$\begin{aligned}\tau \dot{\boldsymbol{w}}(t) &= -\boldsymbol{w}(t) + c y(t) \boldsymbol{x}(t) \\ \tau \dot{w}_0(t) &= -w_0(t) + c_0 y(t) x_0\end{aligned}$$

ここで、 c, c_0 は興奮性、抑制性の荷重値更新に対する学習係数を示す。このモデルは次の2つの実験事実を基にしている：(1) ドーパミンの入力により線条体ニューロンの内部電位が増加あるいは減少する (Hernandez-Lopez et al.,1997), (2) ドーパミン入力により線条体ニューロンのシナプス荷重が変化する (Wickens,1993, Calabresi et al.,1996, Schultz,1998)。つまり、このモデルの自己組織化メカニズムは、強化信号の modulation 下で可塑的に変化する興奮性と抑制性のシナプス荷重の効果のバランスにより調節されることになる。

Caudate ニューロンの response type がこの自己組織化モデルで説明できることを示そう。具体的には、各 1DR の平衡状態 (最終的な安定状態) が4方向の reward 条件でそれぞれの response type を示すことを、またブロックが変わった時に、もともとのブロックの平衡状態を新しいブロックの初期状態とした時に、学習後の平衡状態でそれぞれの response type になることを確かめる必要がある。

まず解析のため、出力関数 f として、step 関数

$$f(u) = \begin{cases} 1 & (u > 0) \\ 0 & (\text{otherwise}) \end{cases}$$

を用いる。

- Q : reward 信号が他の場所 (抑制項や学習の項) に入っていないのはなぜか?
- 中原 : このモデル以外ではだめである理由はない。とりあえずの理由としては、生理実験で projection neuron について、ドーパミンを与えてニューロンの反応をみると、ドーパミンがニューロンの反応に fascilitatory または inhibitory に効いてくる実験がある。

caudate ニューロンの場合、各 reward 方向 (4方向) の入力を $\boldsymbol{x}_i (i=1,2,3,4)$ とする。各 \boldsymbol{x}_i のくる確率を p_i とすると、 p_i は 1DR ブロック内では 1/4 である。

それでは，学習ダイナミクスの解析を始める。確率 p_i で入力 \mathbf{x}_i が与えられ，そのときの強化信号を $\alpha(\mathbf{x}_i)$ とする。各方向の入力で平均した学習則は，

$$\begin{aligned}\tau \dot{\mathbf{w}} &= -\mathbf{w} + c \sum_i p_i y_i \mathbf{x}_i \\ \tau \dot{w}_0 &= -w_0 + c_0 \sum_i p_i y_i x_0\end{aligned}$$

シナプス荷重の平衡状態 $\bar{\mathbf{w}}, \bar{w}_0$ は， $\dot{\mathbf{w}} = \dot{w}_0 = 0$ を満足し，

$$\begin{aligned}\bar{\mathbf{w}} &= c \sum_i p_i y_i \mathbf{x}_i \\ \bar{w}_0 &= c_0 \sum_i p_i y_i x_0\end{aligned}$$

荷重値が収束した後のニューロンの内部状態を \bar{u} とすると，

$$\bar{u} = \bar{\mathbf{w}} \cdot \mathbf{x}(1 + \alpha) - \bar{w}_0 x_0$$

入力 \mathbf{x} に対し， $\bar{u} > 0$ ならば，このニューロンは発火する。 $\bar{u}(\mathbf{x}) > 0$ となる \mathbf{x} の集合を受容野 R と定義する。この R を用いて，

$$p_R = \sum_{\mathbf{x}_i \in R} p_i$$

$$\mathbf{w}_R = \sum_{\mathbf{x}_i \in R} p_i \mathbf{x}_i / p_R$$

とおく。 p_R は全入力のうち R に含まれる入力がかかる確率， \mathbf{w}_R は R の重心を表す。いま，ニューロンの出力関数として step 関数を用いているので，平衡状態でのニューロンの内部状態 \bar{u} は

$$\begin{aligned}\bar{u}_j &= \bar{u}(\mathbf{x}_j) \\ &= c p_R (1 + \alpha_j) \left(\mathbf{w}_R \cdot \mathbf{x}_j - \frac{c_0}{c(1 + \alpha_j)} x_0^2 \right) \\ &= c p_R (1 + \alpha_j) K(\mathbf{x}_j)\end{aligned}$$

と書ける。ここで， $K(\mathbf{x}_j)$ は

$$K(\mathbf{x}_j) \equiv \mathbf{w}_R \cdot \mathbf{x}_j - \frac{\lambda}{1 + \alpha_j}$$

で定義され、 λ は次式で書け、

$$\lambda = \frac{c_0}{c} x_0^2$$

抑制ニューロンの与える影響の大きさを表す。 $K(x_j) > 0$ でニューロンは反応し、 $K(x_j) \leq 0$ で反応しない。ただし、 w_R は各1DRブロックの条件に応じて出てくる値なので、この条件は、必要条件であり十分条件ではない。十分性の確認には、 R 内のすべての x に対し、 $K(x) > 0$ であり、また R 外のすべての x に対し、 $K(x) < 0$ であることをチェックする必要がある。

- Q：抑制性の入力はどこからくるのか？
- 中原：colateral か interneuron のどちらの可能性もある。最近の実験データから、個人的には interneuron の可能性の方が高いと思う。
- Q：interneuron への入力はどこから？
- 中原：最近の実験結果は、大脳皮質からのフィードフォワードの投射が interneuron へいていることを示唆している。
- Q：自己組織化は通常、個々のユニットが同じ特性をもたないように、相互抑制なり、実効的に lateral inhibition を受けるものを入れるが、今のモデルでは？
- 中原：今のモデルではない。単純に言えるのは、caudate ニューロンの出力を全部まとめ step function を通したものを inhibition にするものが考えられる。
- Q：inhibition の効き方は個々のニューロンの可塑性に応じて関わるのか、コンスタントに入っているのかで異なると思うが？
- 中原：検討課題。視床からのフィードバック投射が interneuron に関わってくるという話もある。

caudate ニューロンの解析の前に、入力 x と強化信号 α について述べる。大脳皮質からの入力 x_i は各方向で共通する部分とキュー方向を特徴付ける部分からなり、各方向でオーバーラップしている入力の大さを M 、各方向に選択的な入力の大さを N_i とする。入力の方向優位性は M, N_i の大小で表現される。このことは、大脳皮質からの入力の大さと

各方向の強さを示す実験データを反映している。各方向の入力を単純な形で書くと、

$$\begin{aligned} \mathbf{x}_1 &= (\overbrace{1, \dots, 1}^M, \overbrace{1, \dots, 1}^{N_1}, \overbrace{0, \dots, 0}^{N_2}, \overbrace{0, \dots, 0}^{N_3}, \overbrace{0, \dots, 0}^{N_4}) \\ \mathbf{x}_2 &= (1, \dots, 1, 0, \dots, 0, 1, \dots, 1, 0, \dots, 0, 0, \dots, 0) \\ \mathbf{x}_3 &= (1, \dots, 1, 0, \dots, 0, 0, \dots, 0, 1, \dots, 1, 0, \dots, 0) \\ \mathbf{x}_4 &= (1, \dots, 1, 0, \dots, 0, 0, \dots, 0, 0, \dots, 0, 1, \dots, 1) \end{aligned}$$

このとき、入力間の内積は、

$$\mathbf{x}_i \cdot \mathbf{x}_j = \begin{cases} M + N_i & (i = j) \\ M & (i \neq j) \end{cases}$$

で与えられる。

1DR ブロック中で reward を与える方向を \mathbf{x}_i とすると、この方向に対する強化信号 α_i は $\alpha_i = \alpha$ で、その他の3方向に対しては $\alpha_j = 0 (j \neq i)$ とする。

解析の結果、caudate ニューロンの3つのタイプに応じて条件式が求まる。

a).flexible type

$$M + \frac{N}{2} \leq \lambda < M(1 + \alpha)$$

ここでは、 $N = N_i$ として簡単にして示した。

b).conservative type

1 の方向に空間選択性が高いとし、入力間の特徴を

$$\mathbf{x}_i \cdot \mathbf{x}_j = \begin{cases} M + N_1 & (i = j = 1) \\ M + N & (i = j \neq 1) \\ M & (i \neq j) \end{cases}$$

$$N_1 > N$$

とすると、条件式は

$$M + \frac{N}{2} \leq \lambda < \min\{M(1 + \alpha), M + \frac{N_1}{2}\}$$

c).reverse type

$$(M + \frac{N}{4})(1 + \alpha) \leq \lambda < M$$

ここでも、 $N = N_i$ とする。このタイプでは $\alpha < 0$ である必要がある。

- Q: ここでの α は, TD 誤差か reward そのものなのか?
- 中原: TD 誤差の形式にはなっていない。実験で見られるドーパミンの反応を α とおいている。単純には reward そのものを反映していると思ってもらってもよい。学習の 最初の数回でドーパミンはあるレベルまで達するので, それを表現している。このモデルでは TD 誤差は入っていない。
- Q: このモデルには, 学習の目的を表す objective function はあるのか?
- 中原: objective function はない。安易に objective function は導入するよりも, 豊富な抑制ニューロンとその可塑性, Hebb 学習の自己組織化, 更に強化信号の modulation という, シンプルなモデルからスタートして何がおきるかを見たかった。

さて、先に述べた定理では、ニューロンの出力関数に step 関数を使っていた。(つまり、 $y = f(u)$ の f に、step 関数を使っていた。) この step 関数は、ニューロンが発火しているか、していないかの二値表現の関数だが、これの代わりに、sigmoid 関数を用いると、(正規化した) 平均発火頻度を表現できる。sigmoid 関数ならば、ニューロンの内部状態 u の大きさをニューロンの反応 y の大きさに反映できるからである。これにより、生理実験データにより類似した結果をモデルで得ることができる。なお、ニューロンの刺激方向に関する反応特性は、 M や N_i の各方向の入力の内積を表現している項の大小と対応づけることが可能である。なお、ADR の条件で flexible type の反応を見ると、1DR で報酬が与えられたときと同じ方向への反応を比べると弱くなっている場合があるが、このモデルでも同様な結果が得られる。

まとめると、大脳基底核の線条体ニューロンのモデルとして、強化信号に modulate され自己組織化するニューロンモデルを考えた。このモデルは、線条体内の豊富な抑制性ニューロン、ドーパミン細胞によって計算される強化信号、大脳皮質から基底核への convergent な投射などの基底核の重要な特徴にうまく対応している。数理解析により、実験で得られる caudate ニューロンの異なる反応タイプ (Kawagoe et al., 1998) がモデルのパラメータの違いにより生じることを明らかにした。また、線条体に大脳皮質からの入力の効率のよい表現が得られることを示した。

将来の課題としては、

- 1). 線条体内の抑制性シナプスの可塑性の発見。

2). 大脳皮質から線条体への微少構造との対応関係。

3).reward を時間的な遅れを含む reward に変えた場合について。

などが挙げられる。3) については、モデルの式では α の項を割引き報酬などに変え、数理解析などを行うことはできるが、そのモデルと生理データなどとの対応なども考えないといけない。

質疑応答

- Q: モデルで、細かい実験のタイミング (キュー → action → reward) を導入したらどうなるのか?
- 中原: 興味深い話だと思う。
- Q: 今のモデルでは、シナプスの可塑性ではなく状態変化に reward が入っている。ドーパミン細胞の活動は phasic なので、全体としてドーパミンの影響でニューロンの出力 y が変化するのは一瞬だけだと思うが、それでうまくいくのか?
- 中原: モデルでは、確かにドーパミンは phasic に効くというふうになっている。しかし、一般的にはドーパミンの影響が、神経細胞の内部状態に残り、それ自体は phasic に消える必要はない。なので、このモデルで、特には問題ないと思う。
- Q: flexible type などの条件より実験について suggest できないのか? inhibit の影響や α などが実験的に調べたデータを合ってくるのか?
- 中原: 可能だと思う。 λ は興奮性と抑制性の学習係数の比に抑制性入力の大さの 2 乗をかけたものなので、抑制性入力の大さを変えられるならば、モデルからの予測も可能になる。 λ がある程度低くないと、どのタイプの学習も反応も現れない。なので、例えば、抑制性活動を上げれば、「学習がおきない → どの反応タイプも現れない」という予測はできる。しかし、具体的な実験で実現するには工夫がいる。例えば、muscimol (GABA agonist) を使用すると、projection neuron 自体が GABAergic なので、colateral と interneuron との相互作用の問題がおきてくる。この相互作用の影響を考えると何とも言えない。
- Q: モデルでは、上丘への出力を考えているが、大脳基底核は皮質とループを形成しているのが、大脳皮質への影響もあると思うが。

- 中原：このモデルの動機のひとつは、大脳皮質へのループを還すとき、大脳皮質の入力の表現を効率よく作ることにあり、その理由のひとつが、大脳皮質と線条体のニューロン数が違うところにある。その点では、ループがどのような影響を及ぼすかを調べるのは面白い。時間遅れの reward に対する強化信号の自己組織化を数理的にどうするか面白い問題だと思う。また、この場合皮質との対応関係を考えないといけないと思う。

3 複数の皮質－大脳基底核－視床回路の共同による逐次運動系列の学習，制御

では、もう一つの研究の紹介を始めたいと思います。この研究では、Actor-critic の仮説に乗った上で、大脳皮質－大脳基底核回路が逐次系列運動の学習・制御にどのように貢献しているかということ調べた。そして我々の提案した多重表現仮説に基いてモデルを構築して、彦坂先生の逐次系列運動実験 (Hikosaka et al 1995) での実験結果とシミュレーション結果の対応を調べた (Nakahara 1997; Nakahara, Doya, and Hikosaka, in press)。この研究で着目したのは、一つには、ドーパミンの強化信号 (reinforcement signal) で、もう一つは、大脳皮質－大脳基底核ループは幾つかの分かれたループになっているという解剖学的な特徴だった。Take-home message は cortico-basal ganglia loop (大脳皮質－大脳基底核回路) はパラレルに働き、そのループのそれぞれのサブループは、異なる座標系を利用することによって、新規の系列の学習と良く習熟した系列の遂行を共にうまく実現しているだろうということである。

Hoover and Strick (1993) や Alexander and Crutcher (1990) の解剖学的な研究によって、大脳皮質から大脳基底核には皮質全体から入力が入ってきて視床を介して frontal cortex に出力を主に戻している (なお上丘にも出力を送っている) ということが知られている。そして、実はその大脳皮質－大脳基底核ループは、いくつかの分かれたループを束ねたようになっているということが分ってきた。なお、最近では、小脳の方にもいくつか分かれたループがあるといわれている。例えば、大脳皮質－大脳基底核ループのうちで、motor loop は、運動関連領野、つまり supplementary motor area (SMA), primary motor area (M1), premotor area (PM) などが主に線条体後部 (posterior striatum) に投射してそれがまたループを還している。また、oculomotor loop は、supplementary eye field (SEF),

frontal eye field (FEF) などが posterior caudate の body のあたりに主として投射しループを還している。あるいは prefrontal loop は, dorsolateral prefrontal cortex (DLPF) を中心として posterior parietal も含め線条体前部 (anterior striatum) に投射をしていて, ループを還している。この構造にどんな意味があるのかというのが僕らの研究の出発点になった。同時に思い出してほしいのは, それらが, 非常にリッチな reinforcement signal の modulation を受けているということである。これらが, 生理学的, 解剖学的な知見として研究の出発点となった。

< Serial Button Press Task の図 >

さて彦坂研で行われた実験をまず説明しておこう。この実験課題は逐次系列の課題で, サルがチェンバーの前に座って, 4×4 のボタンがある, そのうちのランダムに2つのボタンが点灯する。それが第1セットで, サルは trial and error でどちらかのボタンを先に押して, 次にもう一つのボタンを押すことを学習しなければならない。もし, 正しい順序で押すことができると, 第2セットに入り, 2つのボタンが点灯する。猿はまた trial and error で正しい順序でボタンを押すことを学習する。もし, 途中で失敗するとまたハイパーセットの最初に戻る。なお, 5セットが1単位のハイパーセットとなっていて, 各々のハイパーセットでは, 5セットは固定でされている。なので, もし第1セットから第5セットまで続けて成功すると, 1トライアル成功したことになる。

< 実験課題のビデオ >

これから見せるのは彦坂研のサルの実験のビデオテープで, 学習初期のサルのパフォーマンスである。ボタンが点灯して, 失敗するとブザーが鳴って第1セットの始めに戻る。何度かくりかえして, サルは一つのハイパーセットを学習していく。1ブロックの実験では, 20回の成功試行をするまで, 同じハイパーセットが続けて提示される。これは非常に巧みな実験課題だと思う。この課題は我々日常の系列学習を抽象的に巧みに表現している。というのは, 実際にサルは一日に例えば20個くらいのハイパーセット (つまり20ブロック) を経験している。このうちの半分はその場限りで経験するハイパーセットで, 残りの半分は毎日同じハイパーセットを経験している。作りうるハイパーセットの組み合わせの数はものすごい数になる (約 7.6×10^{11} 個) ので, このように毎日新しいハイパーセットを経験させつつ, 同時に同じハイパーセットを経験させることができるのである。これを半年とか一年とか, もっとやる。我々の日常生活では, 朝起きて, 歯を磨く, 服を着替える。子供がいる人は分かると思うが, これらは実は学習してできるようになる。つまり, もの

すごい時間をかけて学習をしていって、いまの我々にとっては自動的な行動になっている。一方、例えばある日誰かがテニスを始める、スキーを始める、ワープロを始めるなど、そのときその場で新しい系列課題を学習する。このように、我々の日常生活には、毎日鍛えられて自動的になった系列があるのと同時に、その場その場でこなしていく新しい系列の両方がある。それを非常に抽象的に表現してサルに行わせている。これからさまざまな実験結果とシミュレーション結果をみせるが、その中で、サルが毎日経験して習熟したものを learned hyperset と呼び、その場限りのものを new hyperset と呼ぶことにする。

< learned hyperset の実験のビデオ >

learned hyperset のパフォーマンスをビデオで見せる。これは、日常生活で行っている歯を磨くとか服を着替えることに対応する。実際、一年半くらい課題をやらせないで置いて、久々にやらせても猿はちゃんと覚えていて、本当に植え付けられた系列となっている。

これから、さまざまな実験とシミュレーションの対応を見て行くが、たとえば、線条体前部に GABA agonist である muscimol を注入して、その部分の神経活動を抑えたときには、learned hyperset に対する学習は全然変わらないのに、new hyperset に対する学習は悪くなる。それに対し線条体後部に muscimol を入れたときには習熟した系列, learned hyperset に対するパフォーマンスは悪くなるのに、new hyperset に対するパフォーマンスは悪くならない。このように、様々な機能分化が実験的に見られている。我々は、これから述べるが、多重表現仮説を提案し、その上で、そのモデルのシミュレーション結果と実験結果を比べることによって仮説の妥当性を検討した。

まず、多重表現仮説を手短かに説明しよう。先ほど解剖学的な特徴の話をしたけれど、なんのために、あのような分かれたループがあるのだろうか？機能分化を示している実験は、それぞれのループの違いに対応しているのでは？それぞれのループでの領野の様々な実験での神経細胞の反応の違いとどう関係するのだろうか？これらの問いかけを検討するなかで、各ループでの表現系、言い換えれば、座標系が重要なポイントなのではと考えたのである。

同じ運動系列はさまざまな座標系を使って表現できる。たとえば、同じ系列を視覚座標系で表現することもできるし、腕の関節角度を使うような運動座標系で表現することもできる。それぞれの表現系にはそれぞれの特徴がある。理由はここでは触れないが、特に、視覚座標系の方が学習が早い一方で、運動座標系の方が学習は遅いが、運動座標系で学習

してあれば運動はロバストに遂行できるという特徴の違いが重要だと考えた。先ほどの actor-critic のスキームで考えると、それぞれのループというのは異なる座標系を持ちつつも、ドーパミン細胞からの報酬信号の影響を同様に受けている。つまり、異なる座標系をもった actor に対応する。先ほどの仮説に基づいて考えると、視覚座標の方が学習が早いから、新しい系列を学習するときには、視覚座標を主に使った方が良いというのは想像がつく。その一方で actor として異なる座標系、異なるループで学習を行っていたら、当然お互いの意見に相違が出る。pre-SMA のデータを見ていると、様々なタスクスイッチングのときに発火が多いなどの結果がいろいろある。pre-SMA というのは、visual loop(大脳皮質-大脳基底核の1つのループである DLPF loop をここではそう呼ぶことにする) と motor loop の間であってお互いを繋いでいる。具体的に言うと、皮質では、DLPF と SMA をつないでいるし、基底核の線条体でもこの各々のループの投射部である線条体前部と線条体後部の中間に投射している。この pre-SMA が個々のループのコーディネーションをしているだろうというのが、われわれの仮説のもう一つ重要な点である。

- Q : pre-SMA から SMA に矢印が伸びているが、コーディネートしているという立場から考えると逆向きのほうがわかりやすい感じがする。
- 中原 : 解剖学から言えば、逆向きというよりは両向きのほうが本当はよい。
- Q : どちらの情報も pre-SMA が必要とするように思われるが。
- 中原 : 概念的に考えるといろんなことが考えられる。解剖学的には bilateral なので、もちろん一方向ではない。DLPF と pre-SMA はお互い bilateral に投射しているし、pre-SMA と SMA も bilateral に投射している。その意味で、解剖学的には両方向きの矢印を必要とする。

ここで、一方向にしているには理由があって、一つは pre-SMA のニューロンの反応を見ていると、現在の視覚入力が入ってきて行動を起こすまでの間に反応しているニューロンが結構多い。あとで、また話すが、コーディネーター (pre-SMA) は、DLPF から一部の入力を流すという方法で実現されている。そういう理由でこの矢印はこの向きになっている。ただし、概念的な意味でのコーディネー

タということでは、たしかに他のインプリメンテーションも考えられる。ただ、今のところ生理学的なデータを見てみると今のインプリメンテーションはそんなに嘘ではないし、実験データを十分に反映する良い近似になっているというのが我々の立場である。

<ネットワークモデルの図>

ここでは、大脳皮質だけで示しているが、基本的には我々が考えたのはこのような図の構造のネットワークモデルになる。visual input が DLPF に入ってきて、somatosensory input が SMA に入ってくる。DLPF からある形でコーディネータを介して情報が motor network に伝わる。DLPF では視覚座標系で出力を決めているので、実際に運動を出力するには運動座標系に PMv を介して変換して M1 を介して最終的な運動出力となるという構造を考えている。

それぞれについてもう少し具体的に見ていこう。視覚座標系では二つの機能がある。visual immediate mapping というのは現在の視覚入力に対して出力を計算する。context prediction というのは、一つ前までの情報をもとにして、これから入ってくる入力に対する出力を予測する。visual context prediction であれば視覚座標系、motor context prediction であれば運動座標系でそれぞれ予測を行う。DLPF にもう一つあるといわれているのは作業記憶である。その特徴は、短期的に入力を保持しておいて、必要がなくなればすぐに消去できるところにある。作業記憶というと、短期的に記憶が保持できると言うことが良く注目される。そのアドバンテージは確かに大事だが、その場ですぐキャンセルできるというのも実は重要である。例えば各ブロックで各ハイパーセットの学習を行っているときに、一つのハイパーセットで覚えた入出力表現が次のブロックでは干渉する可能性がある。その意味で、作業記憶でその場で情報を保持して、次のブロックでその情報をキャンセルできるという機能が immediate mapping に入れている。作業記憶をもう少し複雑にすることはできるが、このミニマムモデルでも十分実験との対応がつく。具体的には、immediate visual mapping は、ブロックごとに初期化されて学習する。また、その情報をコーディネータを介して motor context prediction に情報を送る。そして、その情報と motor context prediction の出力との和が取られて、motor network の出力が決められる。この motor network の出力と、PMv を介して送られてきた visual network の出力が M1 で併されて、final motor output が決められ、それによって action が起きる。

visual network では、immediate visual mapping に対する weight に現在の視覚入力 (v^I) がかけられて、visual context prediction に対する

weight には視覚座標に基づくコンテキスト情報 (\mathbf{v}^C) がかけられて、その和に対する softmax によって出力が表現されている。つまり、

$$\mathbf{v}^P(t) = S(\mathbf{W}^{VI}\mathbf{v}^I(t) + \mathbf{W}^{VC}\mathbf{v}^C(t))$$

で与えられる。ただし、 S は softmax function で $S(\mathbf{u})_i = e^{\zeta u_i} / \sum_k e^{\zeta u_k}$ を表す。

motor network では、コーディネータから渡された入力 \mathbf{m}^I を、weight (\mathbf{W}^{MI} ; 実際にはこのモデルでは単位行列である) にかけて、それと運動座標に基づくコンテキスト情報 (\mathbf{m}^C) を motor context prediction の weight (\mathbf{W}^{MC}) にかけて、それらを加えたものに softmax をかけることで出力が得られる。

$$\mathbf{m}^P(t) = S(\mathbf{W}^{MI}\mathbf{m}^I(t) + \mathbf{W}^{MC}\mathbf{m}^C(t))$$

visual network での状態空間は 16 次元で、現在の視覚情報 (\mathbf{v}^I) では、どこに点灯しているボタンがあるかという 0,1 のバイナリ表現になっている。visual context prediction の入力 (\mathbf{v}^C) は、視覚座標系での出力ベクトル \mathbf{v}^O と現在の visual context prediction の状態ベクトル \mathbf{v}^C の差の減衰和で表される。つまり、視覚座標系での過去の行動を時間的に荷重平均したものを示している。

$$\mathbf{v}^C(t+1) = \mathbf{v}^C(t) + \frac{1}{\tau_V}(\mathbf{v}^O(t) - \mathbf{v}^C(t))$$

ただし、 τ_V は減衰の時定数である。motor network の状態表現は、2次元の関節座標上に 64 個の normalized Gaussian を配置し、その activation によって表現される 64 次元のポピュレーションベクトルによって実現される。よってこのモデルでは ill-posed は入っていない。ただし、normalized Gaussian の network が散らしてあるので、多少は ill-posed が入っている。ただし、このモデルでは、visual network の学習が motor network よりも早いというのはパラメータチューニングで入っていて、ill-posed を通じて入っているわけではない。ここで、motor context prediction の出力は運動座標系での出力ベクトル \mathbf{m}^O と現在の motor context prediction の状態ベクトル \mathbf{m}^C の差の減衰和で表される。つまり、運動座標系での過去の行動を時間的に荷重平均したものを示している。

$$\mathbf{m}^C(t+1) = \mathbf{m}^C(t) + \frac{1}{\tau_M}(\mathbf{m}^O(t) - \mathbf{m}^C(t))$$

ただし、 τ_M は減衰の時定数である。学習はこの3つのベクトル ($\mathbf{v}^I, \mathbf{v}^C, \mathbf{m}^C$) に対する weight ($\mathbf{W}^{VI}, \mathbf{W}^{VC}, \mathbf{W}^{MC}$) に関して actor-critic のスキーマで

行っている。最後の motor output の選択は， motor network の出力と visual network の出力を運動座標系になおした後で，積による平均をとっている。

$$p_j(t) = \frac{\mathbf{m}_j^{VP}(t)\mathbf{m}_j^P(t)}{\sum_k \mathbf{m}_k^{VP}(t)\mathbf{m}_k^P(t)} \quad (1)$$

これは情報幾何でいう e-projection になっている。最近，Hinton が products of experts (Hinton 1999) というのをやっていて，似たような定式化のモデルを用いている。

- Q: \mathbf{v} の 16 次元のベクトルはコンテキストに対する重み付けを示しているのか。
- 中原: やっているのは，3つのボックス (immediate visual mapping, visual context prediction, motor context prediction) に対する weight の学習を TD-learning で行っている。normalized Gaussian を介しているので，gradient decent をまじめに取る式になっている。よって， \mathbf{W}^{VI} , \mathbf{W}^{VC} , \mathbf{W}^{MC} のそれぞれに関する更新式は次のようになる。

$$W^{VI}(t+1) = W^{VI}(t) + \eta^{VI} \hat{r}(t) g(\mathbf{v}^O(t), \mathbf{v}^P(t)) \mathbf{v}^O(t) \{\mathbf{v}^I(t)\}^T$$

$$W^{VC}(t+1) = W^{VC}(t) + \eta^{VC} \hat{r}(t) g(\mathbf{v}^O(t), \mathbf{v}^P(t)) \mathbf{v}^O(t) \{\mathbf{v}^C(t)\}^T$$

$$W^{MC}(t+1) = W^{MC}(t) + \eta^{MC} \hat{r}(t) g(\mathbf{m}^O(t), \mathbf{m}^P(t)) \mathbf{m}^O(t) \{\mathbf{m}^C(t)\}^T$$

ただし， η^{VI} , η^{VC} , η^{MC} は学習率， $g(\mathbf{p}, \mathbf{q})$ は softmax のゲイン行列であり，この第 i 対角成分は $(p_i - q_i)q_i(1 - q_i)$ で表され，他の成分は 0 である。これは， $\|\mathbf{p} - \mathbf{q}\|^2$ の自乗誤差に対する stochastic gradient decent を実現している。実際に出てきた系列は，運動座標系で表現された式 (1) の確率分布に基づいて，どのボタンを押すかランダムに選ぶ。また，各ボタンを押すときに毎回ネットワークが回る。

では，シミュレーションと実験の結果の対応を見ていこう。実験で見られたことは，まず short term の学習。1 ブロック内でハイパーセットを学習していくときに，最初に間違いを起こしながら，後半のほうが error が少ない。次に，あるハイパーセットが learned hyperset になっていくときに徐々に学習成果が現れてくるという long term の学習。ということが，実験でも，シミュレーションでも見られた。よって，少なくともそういった形の学習表現になっていると考えられる。

次に，我々の仮説の一つをテストした。思い出して欲しいのは，我々は visual network の方が motor network よりも早く学習するようにパラメー

ターを調整した。このときのやり方は、まず、new hyperset を1つもってきて、それをについて20回のハイパーセットのトライアルを成功するまでにしたエラーの数を指標にして、学習速度をチェックするということを繰り返して、パラメータチューニングを行なった。

しかし、我々の日常はそうではない。つまり、次から次へと new hyperset も来るし、learned hyperset も来る。それでトレーニングシミュレーションの中で、パラメータチューニングした visual network のほうが motor network よりも早いかというのを、visual only と motor only で確認した。すると、十分 visual のほうが早い。そして、この早いのと遅いのをまず単純につなげてみたのが、この no-coordinator である。そのパフォーマンスはおおざっぱにいうと、visual network only と motor network only の中間を示している。早いのと遅いのを足しあわせたので、中間になるのはわかるような気がするが、一般には必ずしもこうなるとは限らない。それを確認したわけである。このままだと visual のパフォーマンスを悪くしただけだが、コーディネーターを入れると、つまり我々のモデルだと、visual よりもパフォーマンスはよくなっている。よって、お互いのコーディネーションの結果が少なくともこのモデルでは現れていることを示唆している。

もう一つモデルのテストをしたのは、working memory をリセットできることが重要であるという仮説で、実際それが単純化された形でモデルには入っている。それをテストとした。どのようにテストしたかという、我々のモデルでは、immediate visual mapping を各ブロック（各ハイパーセット）ごとに初期化している。その初期化をしないでシミュレーションを行った結果がこの黄色のグラフ (no reset) の方で、learned hyperset に対してもパフォーマンスは悪くなっているが、new hyperset に対するパフォーマンスが特に悪くなっていることがわかる。

< no reset の図 >

このようなテストをした上で、実験との比較をした。系列学習の時に大切なのは、その系列をどのように学習しているか、例えば、一つのハイパーセットは一番、二番、三番、四番、五番というつながりになっているが、学習の仕方としては、このセットの中で、このパターンの視覚刺激が来たときにはこの順番という学習の仕方も可能である。そこで、learned hyperset の系列をひっくり返してみる (reverse hyperset)。そのシミュレーション結果が右に示してある。

< reverse hyperset の図 >

この結果が示すことは、reverse hyperset に対するパフォーマンスは、

ほとんど new hyperset に対するパフォーマンスと変わらない。learned hyperset の学習は非常にコンテキスト依存で行われていることがわかる。即ち、系列依存で学習されていることがはっきりわかる。これに関して、追試の実験も行われていて、視覚座標系で reverse の実験を行った結果は、多少パフォーマンスの落ち方がちがう。それぞれ、銅谷さんのチーム (Bapi et al., 2000) あるいは彦坂先生のグループ (Rand et al., 1998, 2000) でテストしている。

次に、opposite hand simulation。これも、我々が行っている逐次系列の学習を考えればわかるが、たとえば、アメリカにいて車に乗って、ウインカーを使おうとして思わずワイパーを回してしまう。つまり、習熟したスキル (系列学習) は、エフェクターに依存していることが多い。あるいはコンテキストに依存している。たとえば、右利きの人が左手でボールを投げるときこちない。この場合、利き腕の問題があるが、まあそういうことを考えるとわかる。実際に実験で個々のハイパーセットを学習させるときには、このハイパーセットは右手でやらせる、このハイパーセットは左手でやらせるという形で実験を行っている。なお、サルは、人間のような利き腕というのではない。そこで、右手で十分学習したハイパーセットを左手でやらせたらどうなるかというのがこの実験の結果である。我々のモデルでは、motor network はエフェクター依存である。

< opposite hand simulation の図 >

この結果が示しているのは、learned hyperset の結果はエフェクター依存なので、手を変えるとそのパフォーマンスは悪くなる。ただし、全く新しい系列に対するパフォーマンスよりは十分良い。エフェクター非依存な記憶を持っているということがこれで示される。

さて、ここから、blockade simulation の話をする。個々の機能の分化を実際にテストした結果を示す。

< blockade simulation の図 >

visual network を blockade している場合、motor network を blockade している場合、coordinator を blockade している場合。それぞれの場合のシミュレーション結果と実験結果を比べるとまあまあ合っているということがわかる。基本的には learned hyperset と new hyperset に対する number of error trials を表現している。白が control で、黄色が blockade。この場合は visual simulation。実験では線条体前部を blockade した場合、まずここで見られるのは、learned, new 両方に対してパフォーマンスが悪くなっている、ただし、new に対するパフォーマンスはより強く悪くなっている。これは後で示す motor network の blockade と比較するとはっきり

する。

< blockade simulation motor network の図 >

motor network の blockade も new と learned を両方悪くする。ただし、visual network と比較すると learned に対するパフォーマンスがより悪くなっている。実験を見ると、線条体後部（我々のモデルでは motor loop にあたる部分）を blockade すると learned のみに対して有意に悪くなる。SMA では new に対してのみ有意に悪くなっている。この意味では、我々のモデルは実験と完全な一対一対応はまだしていない。ただし、我々のモデルではネットワークで blockade しているので、各ネットワーク内の領域の区別はしていない。これはまだ改善の余地があるといえる。

コーディネータの blockade について話そう。これは実験でいうと pre-SMA の blockade にあたるが、両方とも new に対するパフォーマンスのみが有意に悪くなっている。このように、それぞれの blockade でモデルのシミュレーション結果と実験結果の間で非常に良い対応が見つかることが分かった。

< コーディネータの blockade の図 >

では、最後に、dopamine を機能させなくした場合のシミュレーション。motor network(motor loop) に入る dopamine の投射を抑えた場合。あるいは visual network に対する dopamine の投射を抑える場合。このシミュレーション結果は、十分に学習してしまった系列に対しては、dopamine の blockade は visual であっても motor であっても利かない。ただし、new hyperset について言うと、motor network のブロックードは基本的には、visual が学習の初期に利いているので、実際このモデルでは、new に対して motor network を blockade してもパフォーマンスは有意には変わらない。ただし visual を blockade した場合は有意に変わる。これについては実験結果はまだない。これをテストしてみるのはおもしろいのではないかと思う。

< dopamine を機能させなくした場合の図 >

以上まとめると、この研究は、異なるループはどのような機能をしているのかということから出発した。そこで我々は多重表現仮説 (Nakahara 1997; Hikosaka et al., 1999; Nakahara, Doya and Hikosaka, in press) を提案し、その仮説にもとづいて、actor-critic に基づいて、実際にモデルを構築し、 2×5 のさまざまな実験結果との対応を見てきた。するとかなり良い対応が見つかることが示せた。このことは、多重表現仮説の妥当性を示していると考えている。また、異なるループの間のコーディネーションが重要であるということも示せたのではないかと考えている。

質疑応答

- Q: モデルで visual の学習が早い理由は何か。
- 中原: このモデルではパラメータチューニングで実現しているので、その理由自体はモデルの中にはない。理由としては、視覚座標系と運動座標系は一般に ill-posed な関係にあるので、視覚座標系のほうが学習が早いだろうと考えている。というのは、ill-posed というのは、座標系に関する ill-posed もあるが、学習に関する ill-posed もある。学習における ill-posed というのはつまり学習が不安定になってしまう。おそらく座標系における ill-posed, 一対一対応が見つからないということが学習における不安定性引きを起こすので、学習における不安定性を押さえるには運動座標系で十分学習係数を押さえる必要があると推測している。いつかきちんとそこらへんも研究したいと思っている。
- Q: SMA の結果はなぜ合わないと言ったのか。そして始めに言っていたように on-going で研究が進んでいるとうことなので、これからどのようにニューロンモデルなり学習則を変えていこうとしているか。
- 中原: 一つめの質問に関して、今日話したことは、基本的には error したトライアルの回数に関することであったが、実際の実験では performance time とか、実制御の時間に対する実験結果も得られている。SMA の結果を見ると error trial の結果に関しては、実は new に対して悪くなっている。ただ、その一方で、performance time で見ると、学習された系列に対して SMA の blockade が影響しているであろうという結果がある。我々のモデルでは、posterior putamen と SMA を区別せずに learned に対して error trial が特に増えるという結果になっている。そういう意味で、完全には一致していない。二番目の質問に対してだが、たとえば、実時間の話と number of error trials の話をどう組み込んでいくか。それをどのようなモデルでやっていくのがよいのか。もっと単純な数理モデルで攻めるという選択もあるし、あるいは、コーディネータの話で、たとえば丹治先生のグループの帯状皮質運動野の実験結果 (Shima and Tanji 1998) にも対応出来るモデルを考えるなど、やってみるべきことはいろいろある。

- Q：この話では，逐次運動系列課題 (2 × 5 課題) における，visual network による short term の学習と motor network による long term の学習の話題があった。しかし，このモデルでの short term の学習と long term の学習についての基本的な考え方は，運動学習には限らない非常に広い話に還元されるような気がするし，線条体 (striatum) 以外の関係してくるべき部分はたくさんあると思われる。このモデルを外挿していくときに，このモデルの中で最低限含めなければならない拘束条件は何か。
- 中原：運動の実制御の話だと小脳を考慮に入れるべきである。大脳-小脳のいくつか別れたループとの対応がつくのかということも重要である。このモデルを拡張していくときの制約については，どのレベルで考えるかによる。基本的には異なる座標系で共同して強化学習するという意味では外挿していくことは十分可能であるが，cognitive science のコンテキストで外挿していくのか，cognitive neuroscience のレベルで外挿していくのか，もっと cell physiology に近いレベルで外挿していくのかで考えるべき優先課題は変わってくると思う。

以上