

2004年1月22日
独立行政法人 理化学研究所

記憶を使った脳の報酬予測のメカニズムの一端を解明

- 生物の高度知能を実現するロボットの開発に手がかり -

独立行政法人理化学研究所（野依良治理事長）は、脳のドーパミン神経細胞が高度な記憶を使った報酬予測に関わることを、動物実験と計算機シミュレーションを組み合わせることで解明しました。脳科学総合研究センター（甘利俊一センター長）脳数理研究チームの中原裕之研究員、順天堂大学の彦坂興秀教授（現：米国国立衛生研究所、NIH）の研究室と東京工業大学の伊藤秀昭助手の共同研究による成果です。

数ある脳細胞の中で、ドーパミン細胞の機能は近年大変な注目を集めています。例えば、生物の生存に必要な報酬の獲得に関わるとか、複雑な運動を学習するのに役立つことが知られていますし、また、その異常がパーキンソン病などの脳疾患に関わることも知られています。これらのドーパミン（※1）細胞の重要な役割は、その活動が、予測した報酬と実際の報酬がどれくらい違うかといった「報酬予測誤差」を表すことに関係すると考えられます。

しかしながら、実際にドーパミン細胞が、どのような報酬予測誤差を表すのかはまだ良く分かっていませんでした。今回、研究グループでは、動物実験で、適切な手がかりを記憶することで報酬予測が向上するような課題を行わせました。すると、十分課題を経験したあとでは、ドーパミン細胞が、単純な視覚入力のみにも頼る場合よりも、報酬予測誤差をより正確に表すことを発見しました。このドーパミンの活動が、「記憶ありの強化学習」と呼ばれる最先端の計算機学習アルゴリズム（※2）と対応がつくことを示すことに成功しました。

この成果は、教育やスポーツで、適切なアドバイスを得て学習することが重要なことを私たちは日ごろ実感していますが、将来、その脳内メカニズムを明らかにするのに役立ちます。それはまた、高度な学習能力をもつ生物知能を実現するロボットの実現に手がかりを与えます。

本研究成果は、米国の科学雑誌『Neuron』1月22日号に発表されます。

なお、本研究は、文部科学省の特定領域研究（C）・若手（B）の科学研究費助成と岡崎国立共同研究機構生理学研究所による日米科学技術協力事業「脳研究」分野の助成を受けております。

1. 背景

大脳基底核（図1）が、大脳皮質と協力して、複雑な運動の学習・制御に重要な役割を担うことは、以前から知られていました。更に、大脳基底核に豊富に存在するドーパミン神経細胞が、予測した報酬と実際に獲得できた報酬の差を表すような活動を示すことが、近年発見されました。しかも、この活動が、人工知能や知能ロボットで使われる計算機アルゴリズム（※2）（強化学習と呼ばれます。ここで話しているのは、強化学習のアルゴリズムの中で、特に、TD学習、temporal difference learning、と呼ばれるアルゴリズムのことで、以下では、TDモデルと呼ばれてい

る) で用いられる「報酬予測誤差」という(予測した報酬と実際の報酬の違いを示す)学習信号に似ていることが指摘されてきました。この事実は、大変な注目を集めています。何故なら、この学習アルゴリズムは、なかなか強力で、例えば、バックギャモンというゲームでは、このアルゴリズムを使って学習した計算機は、世界レベルの人間のプレーヤーとほぼ同等にプレーすることができますし、現在、ロボットへの応用も進んでいます。つまり、このドーパミン細胞の活動を中心に大脳皮質および大脳基底核回路のメカニズムが解明されれば、生物の優れた学習能力や運動能力の解明に近づくことが期待されますし、同時に、優れた知能ロボットの開発の手がかりにもなると考えられます。

2. 研究手法と成果

このようにドーパミン細胞が報酬予測誤差を表しているらしいということは分かっていますが、実は、その詳細な実態は不明のままでした。例えば、報酬を予測するといっても色々な予測の仕方がありますが、ドーパミン細胞がどのような報酬予測に基づいて誤差を得ているのかも不明でした。我々は、動物実験と計算機シミュレーションを用いることで、ドーパミン細胞が、巧みに記憶を利用した報酬予測に基づいてその誤差信号を出していることを見出しました。

このことを、まず簡単な例を挙げて説明します。あなたが誰かと二人でじゃんけんをする時に勝つ確率(報酬を得る確率)は、普通は $1/2$ ですが、もし相手がパーで引き分けた後には必ずチョキを出すことを知っていれば、勝率は $1/2$ よりも良くなるはずで、このように、うまく適切な手がかりを発見・記憶することができれば、報酬予測の精度を向上することができます。

我々の研究によって、今回、実験課題を十分経験したあとでは、ドーパミン細胞の反応は、適切な手がかりの記憶を必要とする報酬予測誤差を表していることが初めてわかりました。

我々の研究手法とその成果の骨子を具体的に述べましょう。サルが平均すると4回に1回の割合で報酬が得られる課題、つまり報酬の平均確率は $1/4$ である課題を行わせました。ただし、この課題では、実はその前に何回続けて無報酬の回があったか(以下、これを PRN、post-reward trial number と書きます)で報酬の確率が高くなるようになっています(図 2A)。この報酬の確率は、PRN の条件を利用する確率ですので、「報酬の条件付き確率」と呼びます。

実験の初期の頃には、ドーパミン神経細胞の反応を観測してみると、どの PRN の値でも同じ反応をします(図 2B)。これは、単純に報酬の平均確率($1/4$)を利用していると考えられます。一方、後期には、PRN によって異なる反応を示しました。もともとの報酬の条件付き確率と見比べると、報酬確率が低いとき(例 PRN=1、直前の回で報酬を獲得している場合)には、報酬があると大きな反応、つまり“驚いている”わけで、大きな報酬予測誤差を表していることが分かります(図 2B)。このことから、サルは条件付き確率を利用していることが推測されます。

この結果を実証するために、通常の大脳基底核のモデルとされている強化学習のアルゴリズムを用いてこの課題をシミュレーションしてみると、その反応はまったく対応が付きません(図 2C,D 左)。一方、この課題の適切な手がかりである PRN の記憶を利用するアルゴリズムを新たに組んでシミュレーションしてみると、実験

の結果とよく一致しています (図 2C,D 右)。これにより、ドーパミン神経細胞が適切な手がかりを記憶して報酬予測誤差を表していることが分かりました。この結果は、「記憶ありの強化学習」と呼ばれる最先端の計算機学習アルゴリズムとドーパミン細胞の反応の対応がつくことを示しています。

3. 今後の期待

この成果は、近年注目されてきたドーパミン神経細胞の活動を中心に大脳皮質および大脳基底核回路のメカニズムを解明することに貢献します。大脳基底核の回路は複雑な運動系列の学習・制御に重要なみならず、近年、思考をはじめとする認知機能への関与も分かってきています。これらのメカニズムの解明は、教育やスポーツの場面でのアドバイスがどのように学習を向上させるか、適切なアドバイスがどうして大切なのか、などの脳内メカニズムを明らかにするのに役立ちます。この解明は、強化学習などの計算機アルゴリズムの開発・発展と協同で進められるので、高度な学習能力をもつ生物知能を実現するロボットの開発に手がかりを与えます。今後は、大脳基底核や大脳皮質の他の領域との連携を調べて、脳全体として報酬予測がどのように実現されているか調べていく必要があります。

(問い合わせ先)

独立行政法人理化学研究所

脳科学総合研究センター 脳数理研究チーム

研究員 中原 裕之

Tel : 048-467-9663 / Fax : 048-467-9693

(報道担当)

独立行政法人理化学研究所 広報室

Tel : 048-467-9271 / Fax : 048-462-4715

Mail : koho@riken.jp

<補足説明>

※1 ドーパミン(dopamine)

$C_8H_{11}NO_2$ 。神経伝達 (脳の神経細胞間で情報をやりとる際に利用される) 物質の一種。

※2 アルゴリズム

コンピュータを動かして特定の目的を達成させるために必要な情報処理の方法や手順のこと。

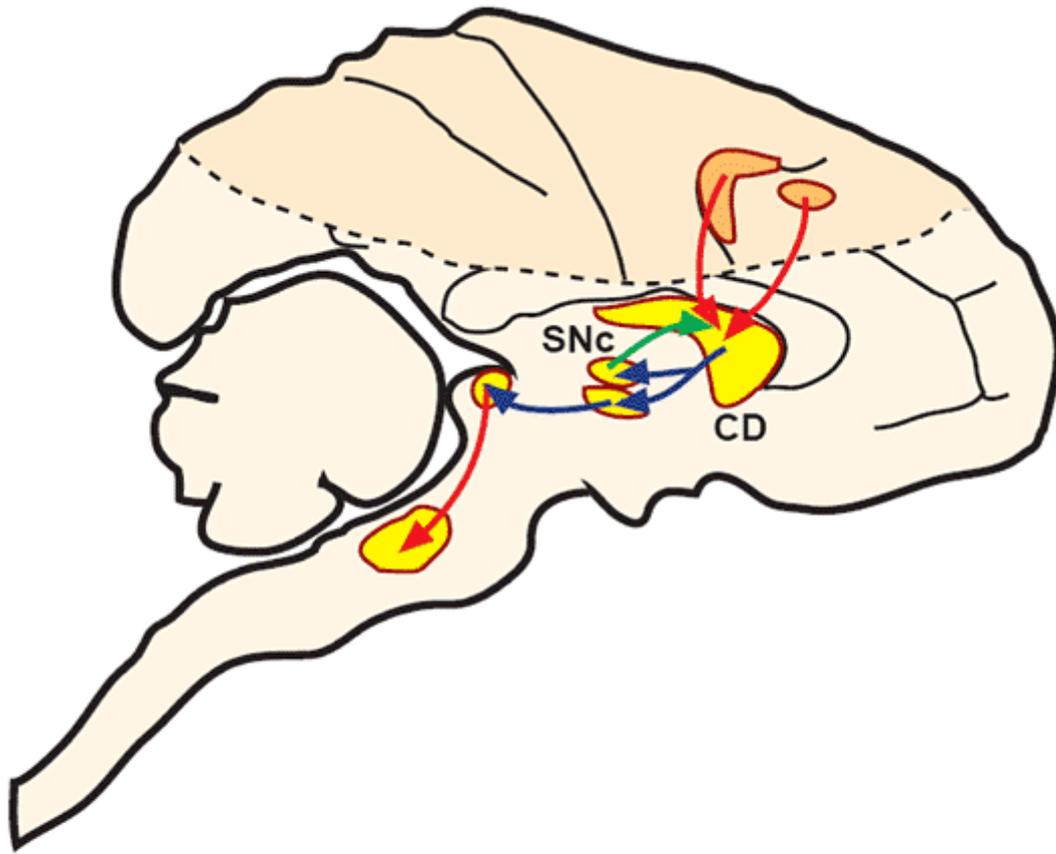


図1 大脳基底核に関する図。ドーパミン神経細胞は、上の図で **SNc** (黒質緻密部) と呼ばれる部位から記録された。**CD** (尾状核) は、大脳基底核の神経核の一つで、その神経活動は、**SNc**に送られている。

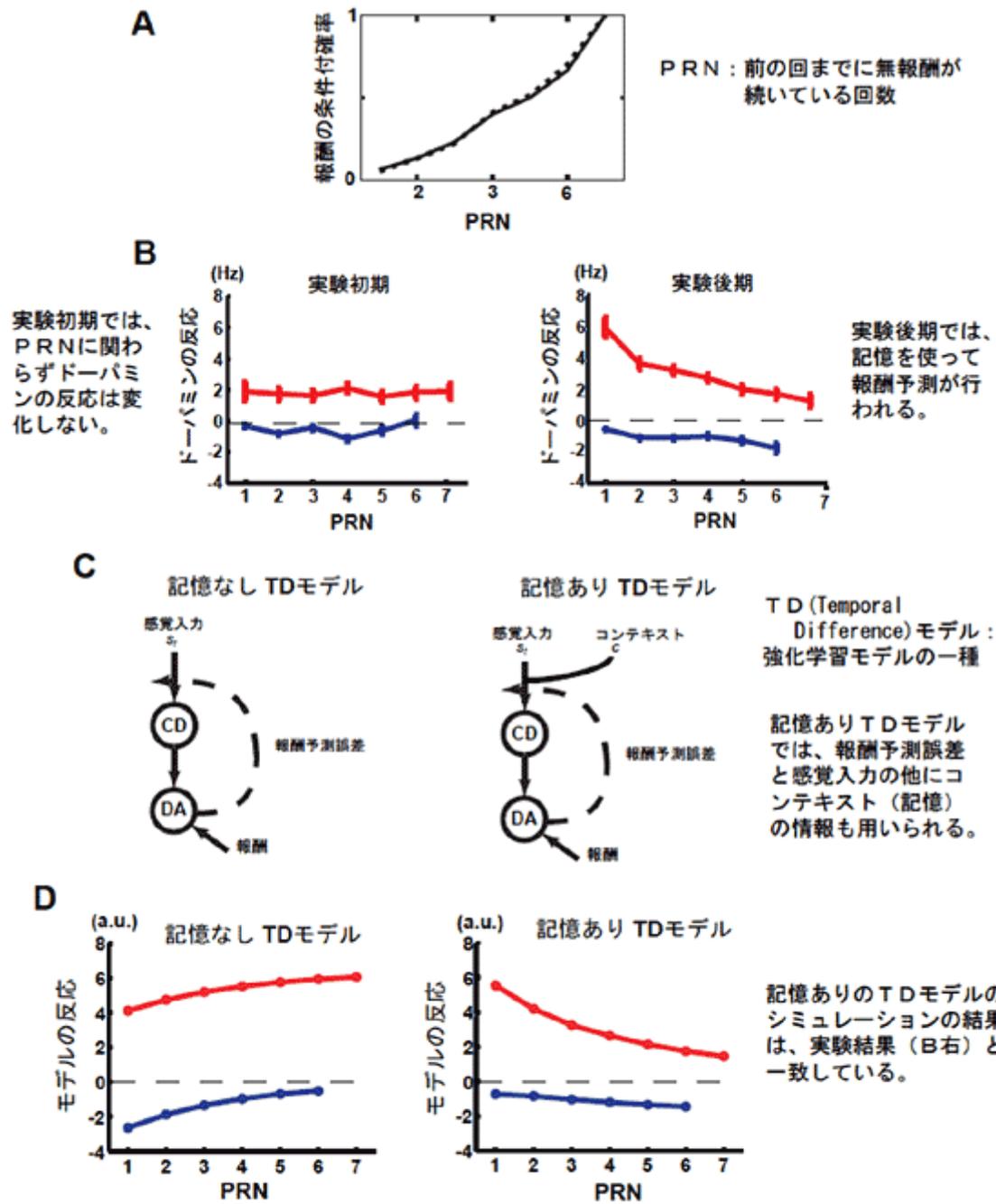


図2 実験結果と計算機シミュレーションの図。B、Dで赤い線は、報酬があった回の反応を示し、青い線は報酬がなかった回の反応を示す。