

## 科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 6 月 10 日現在

機関番号：32639

研究種目：若手研究（B）

研究期間：2011～2012

課題番号：23700503

研究課題名（和文） 中脳ドーパミン系による長期的な将来報酬価値の脳内計算機構

研究課題名（英文） Long-term reward value coding by midbrain dopamine neurons

研究代表者

榎本 一紀（ENOMOTO KAZUKI）

玉川大学・脳科学研究所・嘱託研究員

研究者番号：10585904

研究成果の概要（和文）：ニホンザルを被験体とした神経生理学実験と、強化学習モデルを用いた計算理論によるシミュレーション・データ解析を組み合わせることによって、複数の報酬獲得ステップを経てゴールに到達する状況において、中脳ドーパミン細胞が、将来期待される複数の報酬情報を表現することを明らかにした。また、このような情報表現は、学習によってはじめて獲得されることを示した。これらの成果は、遠い目的に到達するための意思決定や行動選択にかかわる神経回路メカニズムの解明に貢献する。

研究成果の概要（英文）：We showed that midbrain dopamine neurons encode long-term value and its expectation errors of future rewards. By using electrophysiological methods and computational simulations based on reinforcement learning theories, we recorded and analyzed dopamine neuron responses in monkeys performing a behavioral task in which they obtained multiple rewards by multi-step actions. We also confirmed that dopamine neurons evolved these activities by learning the task. These functions play an important role in decision-making and action selection to achieve a distant goal.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
交付決定額	2,900,000	870,000	3,770,000

研究分野：総合領域

科研費の分科・細目：脳神経科学・融合社会脳科学

キーワード：ドーパミン・霊長類・報酬価値・予測誤差・意思決定・強化学習・大脳基底核

## 1. 研究開始当初の背景

われわれ人間や他の動物にとって、目先の損得だけでなく、更に先の将来にわたる長期的な報酬を予測し、それに基づいた意思決定・行動選択を行うことは、コストを最小化し、利益を最大化するために必須であり、生き残るために重要である。そのような長期的な将来報酬の情報を表現している脳神経メカニズムの解明は、神経科学のみならず、心理学や経済学などの多くの分野からも期待されている。大脳皮質-基底核系は、報酬に基づいた行動の発現に中心的な役割を果たしていることが知られている神経システム

であるが、そのなかで中脳ドーパミン細胞は、大脳前頭葉皮質や辺縁系、線条体などの広範囲にわたる脳部位に情報を送り、それぞれの脳部位でのシグナル伝達の可塑性を調節している。現在までの研究によって、ドーパミン細胞は食べ物や金銭などの報酬に対して放電応答を示し、その放電頻度の変化は、刺激の新規性や報酬の期待値を表現することが知られている（Salamon et al, 2005）。条件刺激によって報酬が予告されていると、条件刺激に対して予測される報酬の価値（報酬量×報酬確率）を反映した応答がみられ（Schultz et al, 1997, Morris et al, 2004）、

実際に報酬が得られた/得られなかったときには、予測との誤差を反映した応答を示す (Satoh, Kimura et al, 2003)。そのような細胞活動は強化学習モデル (Sutton and Barto, 1998) でよく記述できることが知られている。強化学習モデルのうち代表的なものは TD 学習 (時間的差分学習) であり、報酬価値に基づいた意思決定と行動選択を記述する計算論的モデルとして、動物の神経システムが実装していると考えられている。この学習モデルでは、行動選択のための価値の推定に、ゴールに至るまでの各ステップで得られることが期待される報酬の積算値を、より遠いステップのものほど価値を割り引いて (割引率:  $0 \leq \gamma \leq 1$ ) 利用する。しかしながら、従来までの多くの研究においては、報酬価値・予測誤差を推定するにあたって、1ステップで得られる1回分の報酬価値についてしか調べられておらず、動物が長期的な将来報酬を見込んで価値判断をしている際に、その報酬価値がどのようにドーパミン細胞および線条体細胞の活動を調節するのかについては明らかではない。

## 2. 研究の目的

ドーパミン細胞の活動が TD 誤差を表現するのであれば、それは、上述したような、長期的な将来報酬情報を反映しうはずである。つまり、現在の報酬だけでなく、2手先、3手先でも報酬獲得が予測できる状況であれば、それらも含めた、累積的な報酬情報が細胞活動に反映されると考えられる。この仮説を検証するため、ニホンザルに複数ステップの報酬獲得を経てゴールに到達する行動課題を学習させ、課題遂行中の細胞活動を記録し、強化学習モデルを用いた計算機シミュレーション・データ解析を行うことで、ドーパミン細胞および線条体細胞が長期的な将来報酬価値を表現するかどうかを検証する。

## 3. 研究の方法

ニホンザルを被験体として、1ブロックが3回の報酬獲得試行からなる行動課題を学習させ、課題遂行時の細胞活動を電極記録する (図1)。この課題では、3つの選択肢 (押しボタン) のうち1つが報酬を得られる正解ボタンとなっており、動物は試行錯誤で正解ボタンを探す。正解ボタンを選んだ時は、報酬として口元のパイプからジュースが与えられる。最初の試行 (第1探索試行) で正解ボタン以外を選び、報酬が得られなかった場合は、次の試行 (第2探索試行) で前試行とは別のボタンを選び、そこでも正解できなければ、更に次の試行 (第3探索試行) で、残った1つのボタンを選ぶことになる。いずれの場合でも、正解すれば、続く2回の試行でも同じボタンを選択することで、2回目、3回

目の報酬が得られる (第1繰り返し試行・第2繰り返し試行)。第1探索試行から第3探索試行にかけて報酬確率 (報酬獲得率) は段階的に上がってゆき、第1繰り返し試行・第2繰り返し試行においては、直前の行動を繰り返すだけであるので、ほぼ確実に報酬が得られる。よって、もし動物が現在の報酬のことにしか考えていないのであれば、各試行開始時に予測される報酬の価値は試行ごとの報酬確率と相関するであろうが ( $\gamma=0$ )、1ブロック全体の報酬を見込んでいるのであれば、第1~第3探索試行開始時から、第1・第2繰り返し試行の報酬も含んだ3回分の報酬を予測しているはずである ( $0 < \gamma$ )。

### ターゲット選択

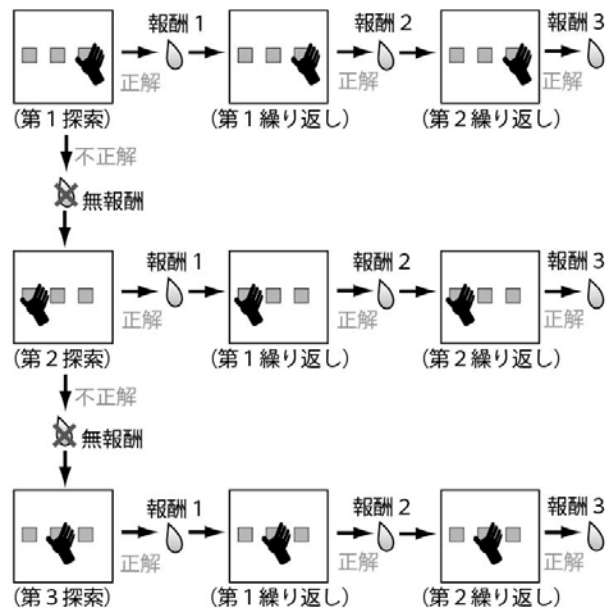


図1 動物が学習した行動課題

ドーパミン細胞は課題開始の合図である視覚刺激と強化因子に応答し、それぞれの応答は報酬価値と予測誤差を表現することが分かっているので、強化学習則に基づいた計算機シミュレーションによって算出した価値と実際の応答を比較することで、将来報酬の割引率などの定量的な解析を行う。実験終了後、脳標本を作製し、細胞の応答特性と記録部位を検討することで神経回路レベルの脳機能に迫る。

## 4. 研究成果

3頭のニホンザルから、185個のドーパミン細胞の活動を記録し、課題遂行時の応答を調べた。ドーパミン細胞は各試行開始時の視覚刺激 (条件刺激) に対して一過性の興奮応答を示した (図2A)。放電頻度の大きさは、第1探索試行から第3探索試行にかけて、報酬確率に伴って増大したが、ほぼ確実に報酬が得られる第1・第2繰り返し試行では、むしろ第3探索試行よりも減少していた (図2B)。

このことは、第1～第3探索試行開始時には、第1・第2繰り返し試行の報酬も含んだ3回分の報酬価値を、第1繰り返し試行開始時には、第2繰り返し試行の報酬も含んだ2回分、第2繰り返し試行開始時にはその試行1回分だけの報酬価値を（時間的に遠い報酬ほど価値を割り引いて）それぞれ予測しており、そのような複数の報酬情報が細胞応答に反映されていることを示唆している。これらの応答をよく予測できるような将来報酬の割引率（ $\gamma$ ）を計算機シミュレーションによって求めたところ、その値は動物の行動データ（強化因子提示前のパイプなめ行動）から算出した値とよく一致していた。また、予測誤差を表現すると考えられている、行動選択の結果（正解／不正解）を知らせる音刺激（強化因子）に対する応答は、同じ割引率の値を用いて計算した、複数回の報酬も含んだ TD 誤差の値との間に有意な相関が認められた（図 2C）。

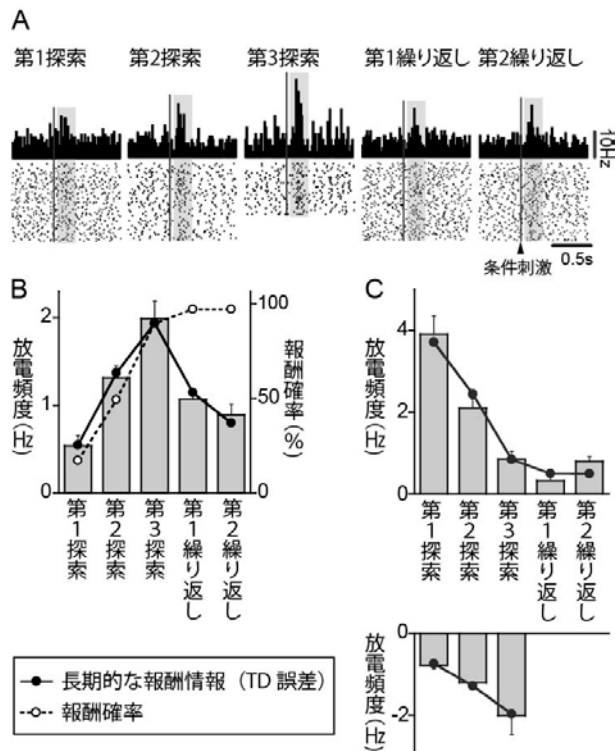


図 2 (A) 代表的な単一ドーパミン細胞の条件刺激に対する応答。試行ごとにヒストグラム（上）とラスタプロット（下）を表示している。(B) 課題習熟後における、条件刺激に対する 26 個のドーパミン細胞集団の平均応答（棒グラフ）。折れ線グラフは、強化学習理論に基づいて推定した、長期的な報酬価値（実線）と各試行での報酬確率（点線）。(C) (B) で示したドーパミン細胞集団の強化因子に対する平均応答（棒グラフ）。正解時（上）と不正解時（下）に分けて表示している。折れ線グラフは (B) で求めた割引率（ $\gamma=0.31$ ）を用いて計算した報酬予測誤差。

このような細胞活動は一朝一夕にあらわれるものではなく、3 回の報酬が得られるという課題構造を学習することによって獲得されると考えられる。そこで、この行動課題を学習している途中の時点と、課題習熟後の時点でのドーパミン細胞の活動を比較した。日ごとの報酬獲得率の上昇を学習の進み具合の指標として、学習期間を初期と後期に分けて応答を調べたところ、条件刺激に対する応答は、学習初期には小さく、放電頻度に試行間の差が認められなかったが、学習後期になって課題構造に習熟すると、ドーパミン細胞は、第3探索試行の応答が最大になるような、大きな放電を示すようになった。それぞれの応答から計算した割引率（ $\gamma$ ）の値は、学習初期で0.04、学習後期で0.38となった。つまり、学習初期では次試行以降の報酬価値をほぼゼロまで割り引いているのに対して、学習後期では、次試行、次々試行の報酬価値も含めた、長期的な報酬情報を反映した応答が見られるようになったといえる。以上の結果から、学習によって、ドーパミン細胞の活動が長期的な将来報酬情報を表現することが確かめられた。

以上の結果を論文にまとめ、発表した。さらに、山田洋（国立精神・神経医療研究センター）と共同で行った研究においては、同じ行動課題を用いてニホンザルの線条体から神経細胞活動記録を行い、ドーパミン細胞が情報を送っている先の線条体においても、運動情報などとともに、長期的な将来報酬の情報が表現されていることを確かめ、論文として発表した。また、ドーパミン細胞活動の解析を進め、中脳において、長期的な報酬情報の表現に部位特異性が認められる結果を得たので、これらの成果をまとめ、学会や研究会で発表した。

これらの結果は、人間や他の動物が、変動しつづける環境の中で、試行錯誤によって利得を最大化していくメカニズムとして、神経システムに実装されていると考えられている強化学習モデルについて、物質的基盤となる知見を提供する。また、目先の利得にとらわれず、長期的なゴールに向かって意思決定・行動選択をするための神経システムとして、ドーパミンや大脳基底核が中心的な役割を果たしていることを明らかにし、詳細な神経回路の解明に貢献する。

## 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 2 件）

(1) Yamada H, Inokawa H, Matsumoto N, Ueda Y, Enomoto K, Kimura M. Coding of the long-term value of multiple future rewards

in the primate striatum. *J Neurophysiol.* 109(4):1140-1151 (2013). 査読有  
DOI: 10.1152/jn.00289.2012

(2) Enomoto K, Matsumoto N, Nakai S, Satoh T, Sato TK, Ueda Y, Inokawa H, Haruno M, Kimura M. Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proceedings of the National Academy of Sciences* 108(37):15462-15467 (2011). 査読有  
DOI: 10.1073/pnas.1014457108

〔学会発表〕 (計4件)

(1) Enomoto K, Matsumoto N, Kimura M. Location-specific roles of dopamine neurons in the midbrain in learning and decision-making. *Neuroscience 2012 (SfN's 42nd annual meeting)*. 2012.10.13. Ernest N. Morial Convention Center, New Orleans, USA

(2) 榎本一紀. ドーパミン神経系による予測と意思決定の制御機構 (霊長類における研究を中心に). 第34回日本生物学的精神医学会. 2012.9.28. 神戸国際会議場, 兵庫県

(3) 榎本一紀, 松本直幸, 木村實. 学習と意思決定課題におけるサルの中脳部位特異的なドーパミン細胞の活動 Heterogeneous signals of midbrain dopamine neurons during learning and decision-making. 第35回日本神経科学大会. 2012.9.18. 名古屋国際会議場, 愛知県

(4) 榎本一紀, 松本直幸, 木村實. 中脳ドーパミン細胞による、異なった時間スケールでの報酬価値とその誤差の表現 Midbrain dopamine neurons encode reward value and its error on heterogeneous time scale. 第34回日本神経科学大会. 2011.9.16. パシフィコ横浜, 神奈川県

〔図書〕 (計1件)

(1) 榎本一紀, 木村實. 羊土社. 実験医学増刊 Vol.30 No.2 「in vivo 実験医学によるヒト疾患解明の最前線」. 2012. 257

〔その他〕

ホームページ等

(1)<http://physiology.jp/exec/page/stopics76>

(2)[http://brainprogram.mext.go.jp/media/press/110905tamagawa\\_kimura\\_press2.pdf](http://brainprogram.mext.go.jp/media/press/110905tamagawa_kimura_press2.pdf)

6. 研究組織

(1)研究代表者

榎本 一紀 (ENOMOTO KAZUKI)  
玉川大学・脳科学研究所・嘱託研究員  
研究者番号:10585904

(2)研究分担者

なし

(3)連携研究者

なし