

## 特集 精神疾患の病態研究の最前線

## 遅延報酬の割引に対するセロトニンの効果 ——精神疾患の病態理解への応用——

岡本 泰昌<sup>1)</sup>, 岡田 剛<sup>1)</sup>, 志々田 一宏<sup>1)</sup>, 福本 拓治<sup>1)</sup>, 町野 彰彦<sup>1)</sup>,  
山下 英尚<sup>1)</sup>, 田中 沙織<sup>2)</sup>, 銅谷 賢治<sup>3)</sup>, 山脇 成人<sup>1)</sup>

われわれは、日常生活において、報酬の多寡と報酬が得られるまでの遅延時間を勘案し様々な行動選択を行っている。遅延報酬割引は、後で得られる報酬の価値を決定する理論的神経科学の概念である。遅延報酬割引が急峻な場合には、目先の報酬をより選択することで定義される衝動的な選択がおこる。われわれの行動学および神経画像的データは、ヒトにおいて大脳皮質-基底核回路における線条体の機能修飾を介して、セロトニンが遅延報酬割引率を調整することを明らかにした。すなわち、われわれが提示したセロトニンの働きは、選択的セロトニン再取り込み阻害薬が有効とされるうつ病、パニック障害、強迫性障害といった精神障害でみられる衝動性の一面を示しているとも考えられる。これらの所見を確認するためには、遅延報酬パラダイムを用いたこれらの精神障害を対象とした検証が必要である。

<索引用語：精神疾患，遅延割引，報酬，線条体，セロトニン>

### 1. はじめに

「人気のレストランの行列に並ぶか、隣の普通の店で空腹を満たすか」「渋滞でいつ来るかわからないバスを待ち続けるか、歩いて駅に向かうか」など、今すぐに目的や欲求が満たされないときに、どれだけ辛抱強く待ち続けるか、あきらめて別の手段を取るかは日常生活でよく直面する問題である。すなわち、われわれは報酬の多寡と報酬が得られるまでの遅延時間を勘案し様々な行動選択を行っている。そこで、本稿では、「遅延報酬の割引に対するセロトニンの効果——精神疾患の病態理解への応用——」と題して、その理論的背景として強化学習理論、動物実験の結果からこれまで明らかになっている遅延報酬の割引率とセロトニンの関連性について述べた上で、われわれがヒトを対象として行ってきた脳機能画像研究や

行動実験の結果を紹介し、これらの研究結果の精神疾患への応用の可能性についてまとめる。

### 2. 強化学習理論

動物行動学や心理学の分野では、えさや水の獲得は報酬、痛みなどは罰あるいは負の報酬であり、ある行動の結果として、報酬が得られる、あるいは罰が回避できると、その行動が「強化」されるという原理で行動の学習過程が説明されている<sup>1)</sup>。SuttonとBarto<sup>15)</sup>は、行動心理学の「強化」の概念をヒントに、探索的な行動学習の理論として「強化学習」を1980年代に定式化した。その後、強化学習は最適制御理論との融合により理論的な解析とロボット制御などへの応用が進められた。そして今日、さらに強化学習の理論は、人間や動物の行動学習の脳機構に関しても、その神経回路

著者所属：1) 広島大学大学院精神神経医科学

2) 大阪大学社会経済研究所

3) 沖縄科学技術大学院大学神経計算ユニット

と物質系の複雑な振る舞いを理解するためのモデルとして認識されるに至っている。

すなわち、強化学習は、行動の良し悪しを評価する報酬信号をもとに、長い目で見てより多くの報酬を得られるような行動則を探索的に学習する理論的枠組みである。その特徴は、将来的にどれくらいの報酬が得られるかを予測する「報酬予測」と、その予測値の変化を元に行動を変化させる「行動強化」という2つの学習が、ともに報酬の予測誤差から与えられる「TD 誤差」をもとに行われるという点である。

時刻  $t$ 、状態  $s$  での報酬予測  $V$  は次のように定義される。

$$V(s(t)) = E[r(t+1) + \gamma r(t+2) + \gamma^2 r(t+3) + \dots]$$

( $E$ : 期待値,  $r$ : 報酬,  $\gamma$ : 割引定数)

このとき、将来得られる報酬の価値は遠い将来のものほど割引いて評価されると仮定し、式には割引定数  $\gamma$  ( $0 < \gamma < 1$ ) が含まれている。つまり報酬を予測し、将来得られる報酬の総額を最大にするような行動ルールを学習するという問題となる。またこの問題は、「現在の予測」と「1歩前の予測」との差である報酬予測誤差 (TD 誤差)

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

がゼロになるような行動ルールを学習するという問題になる。

ドパミンニューロンが報酬予測誤差を表現していることを示した Schultz らの研究をはじめ<sup>10)</sup>、脳活動イメージングを用いて大脳基底核の一部である線条体が報酬予測誤差に関わることを示した O'Doherty らの研究や<sup>11)</sup>、数理モデルに基づく多くの先行研究によって、報酬に基づく行動学習や行動選択の脳のメカニズムが明らかになりつつある。

### 3. 遅延報酬の割引率とセロトニン

上記の報酬予測の式において、将来の報酬の価値をどのくらい速く割引いて評価するかを制御しているのが、割引定数  $\gamma$  である。 $\gamma$  の設定により目先の報酬にとらわれた行動や、より長期的に

見て有利な行動が選択される。 $\gamma$  が小さいと将来の報酬に関する価値が急激に割引かれ、近視眼的な報酬に基づく行動を取るようになる。それに対して、 $\gamma$  が大きいと将来の報酬に関する価値は緩やかに割引かれるため、長期的な報酬予測が可能となると考えられる。つまり遅延報酬価値の割引定数は、どれくらい先の報酬まで考慮に入れて行動を選択するかという、報酬予測の時間スケールを決めるパラメータである (図 1)。

われわれは日常的に、周囲の状況や現在の行動から、即座に得られる結果と長期的な結果の双方の予測をもとに行動を選択している。利益や損失、快楽や苦しみなどの「報酬」は、行動の結果直ちに得られるものと将来的に遅れをもって与えられるものがあるが、その双方を正しく予測し、その適切なバランスをもとに行動を選ぶことは人間の知的機能にとって非常に重要である。例えば、将来的に大きな報酬が得られる行動よりも、即時的に少ない報酬を得られる行動を頻繁に選んでしまう「衝動的選択」は、短期と長期の報酬予測のバランスが崩れることと考えることができる。

このような衝動的選択を引き起こす損傷部位としては、前頭葉 (内側部, 眼窩面皮質), 側坐核などが先行研究において報告されている<sup>3,10)</sup>。また神経伝達物質のひとつであるセロトニンと衝動的選択の関係が、多くの先行研究から報告されている<sup>2,9,20)</sup>。このように衝動的選択に関わる脳部位や物質は明らかになりつつあるが、それらがどのようなメカニズムで衝動的選択を引き起こすのか、また脳部位と神経伝達物質系の関係などいまだ不明な点が多い。

そこでわれわれはこれらの点を検証するために以下に示す検討を行った。

### 4. 異なる時間スケールでの報酬予測に関わる脳機構

まず、異なる時間スケールでの報酬予測に関わる脳機構を明らかにするために、長期的な報酬の予測と短期的な報酬の予測がそれぞれ必要な行動学習課題を新たに考案し、ヒトが予測を行って

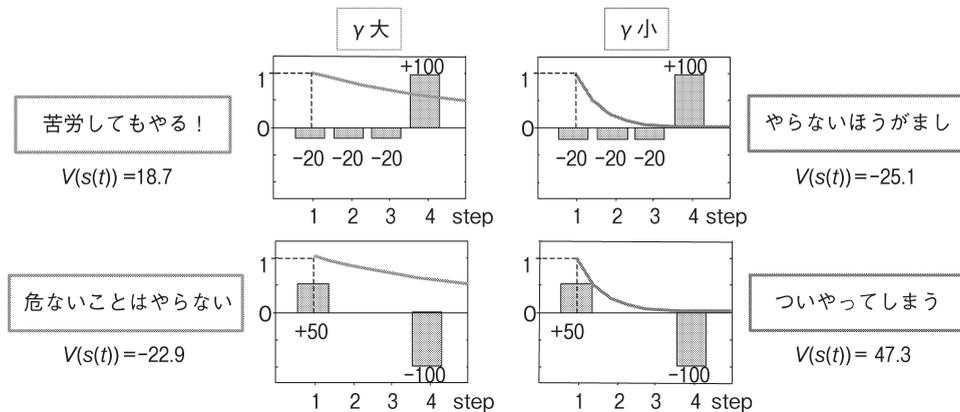


図1 報酬予測の時間スケール

価値関数  $V$  は、将来得られると予測される報酬の総量で、ある時刻  $t$ 、状態  $s$  において  $V(s(t)) = E[r(t+1) + \gamma r(t+2) + \gamma^2 r(t+3) + \dots]$  ( $E$ : 期待値,  $r$ : 報酬) と定義される。一方、割引定数  $\gamma$  ( $0 < \gamma < 1$ ) は、どれぐらい先の報酬まで予測するかを示し、 $\gamma$  が大きければ遠い将来に得られる報酬を考慮した評価が可能となり、 $\gamma$  が小さければすぐに得られる報酬のみを考慮した評価となる。すなわち、 $\gamma$  が大きければ目先の損失があっても、将来の報酬が見通せれば、 $V(s(t))$  はプラスとなり「苦労してでもやる」であろうし、目先の報酬があっても、将来の損失が見通せれば、 $V(s(t))$  はマイナスとなり、「危ないことはやらない」という選択になる。逆に  $\gamma$  が小さければ将来の報酬があっても、目先の損失のみしかみえず、 $V(s(t))$  はマイナスとなり「やらないほうがまし」であろうし、将来の損失があっても、目先の報酬のみしかみえず、 $V(s(t))$  はプラスとなり、「ついやってしまう」という選択になる。

るときの脳活動を、機能的磁気共鳴画像法 (fMRI) を用いて測定した<sup>17)</sup>。

20人の被験者は、MRI装置の中でボタン押し課題を行った。画面上に提示される3種類の図形に対して、左右それぞれのボタンに応じた報酬と、次に表示される図形が決定される。被験者はこのルールを試行錯誤によって学習し、最終的により多くの報酬を得られるような行動を取得する。短期報酬予測条件では、被験者は単純に各図形に対して、より多くの報酬金額を与えるボタンを押すことを学習する。一方、長期報酬予測条件において大きな正の報酬が得られる図形を呼び出すには、まず小さな負の報酬を受けるボタンを選ばねばならない。つまり、目先の報酬にとらわれていては、長い目で見て最適な行動を取ることができない。この2つの条件で被験者に交互に学習を行ってもらい、その脳活動を比較した。

その結果、短期報酬予測条件では前頭葉の下部や大脳基底核の一部に、長期報酬予測条件では前

頭葉の外側部や頭頂葉、大脳基底核、小脳、また脳幹でセロトニンを伝達する細胞を多く含む縫線核に活動の増加が見られた。

次に、脳の各部位がどのような時間スケールにおける報酬予測に関わるかを調べるために、被験者の脳活動データを強化学習理論に基づいて解析した(図2)。その結果、前頭葉と側頭葉の間に位置する島皮質の腹側部から背側部に向けて、短い時間スケールから長い時間スケールでの報酬予測値に相関する脳活動のマップを明らかにした。また大脳基底核の入力部にあたる線条体では、その腹側部から背側部に向けて、短い時間スケールから長い時間スケールでの報酬予測誤差に相関する活動のマップを明らかにした。この結果は、島皮質と線条体の間に部位対応連絡が存在するという解剖学的所見とも一致していた。今回の実験結果は、これまで情動的な機能を司るとされていた線条体腹側部を含むネットワークが短期的な報酬予測に関わり、より高次な認知的機能を司るとき

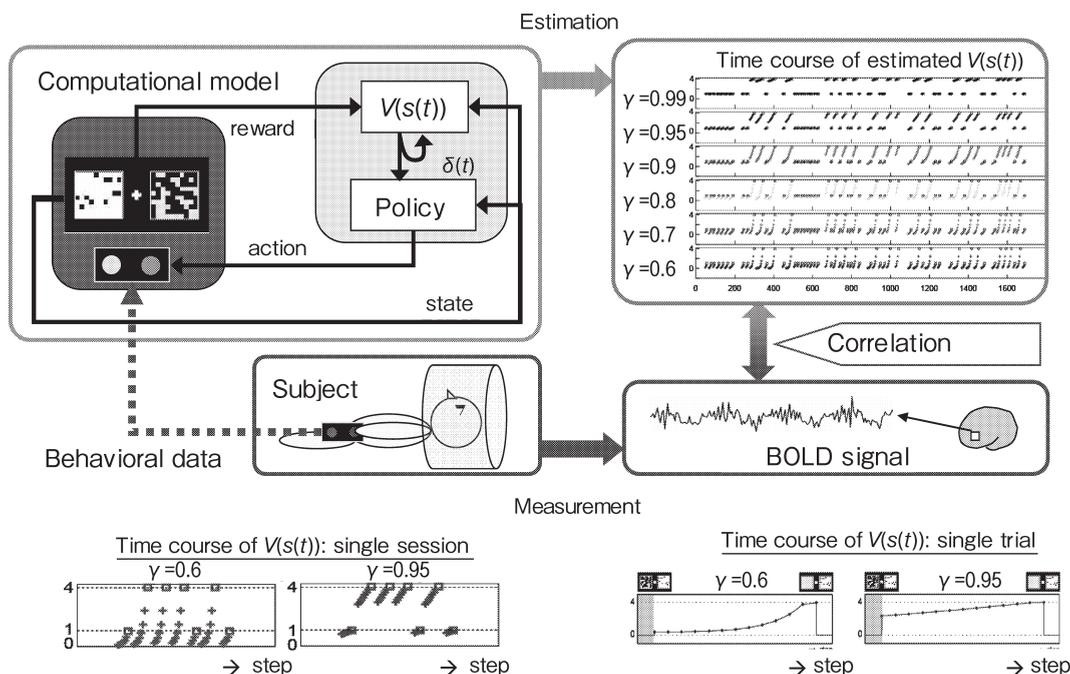


図2 強化学習理論モデルによる脳活動データ解析

各被験者が実際に経験した図形と報酬の時系列を、強化学習のプログラムに疑似体験させ、報酬予測の学習を行わせた。この際、異なる時間スケールの報酬予測は脳の異なるネットワークが関わるという仮説に基づき、予測の時間スケールを決める割引定数  $\gamma$  を6通りに変えて学習を行わせた。このようにして学習プログラムが推定した各被験者にとっての報酬の予測値  $V(s(t))$  と予測誤差信号 ( $\delta$ ) と、各被験者の脳活動データとの相関を調べた。この解析において、それぞれの信号と高い相関の見られた部位は、その信号の計算に関わる可能性が高いことを示している。

下段右側は、1回のトライアル中の  $V(s(t))$  の経時変化を表し、下段左側はセッション中の複数のトライアルの  $V(s(t))$  の経時変化を表している。

れてきた線条体背側部を含むネットワークが長期的な報酬予測に関わるという、並列ネットワークの時間スケールでの機能分化を示唆している。

### 5. セロトニンが線条体を通る

#### 並列ネットワークの活動を制御する

異なる時間スケールでの報酬予測を行う並列ネットワークへのセロトニンの影響を調べるために、セロトニンの前駆物質であるトリプトファンの経口摂取により、被験者のセロトニンレベルを人為的に調整した状態で、報酬予測課題を行わせ、そのときの行動と脳活動を測定した<sup>18)</sup>。

12人の被験者は、セロトニン前駆物質のトリプトファンが不足、通常、過剰な状態のアミノ酸

混合飲料を1日に一種類飲み、3日間にわたりそれぞれのセロトニンレベルで実験タスクを行った。この実験は、被験者のみならず実験者も、その被験者がその日どの条件か知らないという二重盲検法によって行われた。被験者はすぐに得られる小さい報酬と、時間はかかるが大きい報酬の選択課題を行った。

1試行中の報酬予測に関する脳活動を捉えるために、先の実験と同様に強化学習に基づいた脳活動データ解析を行った。被験者の実際の行動データから推定した報酬予測とfMRIデータの相関を調べた。コントロール状態では線条体に、腹側部から背側部にかけて短期から長期の報酬予測に関わる活動が見られた。この結果は、先の実験の

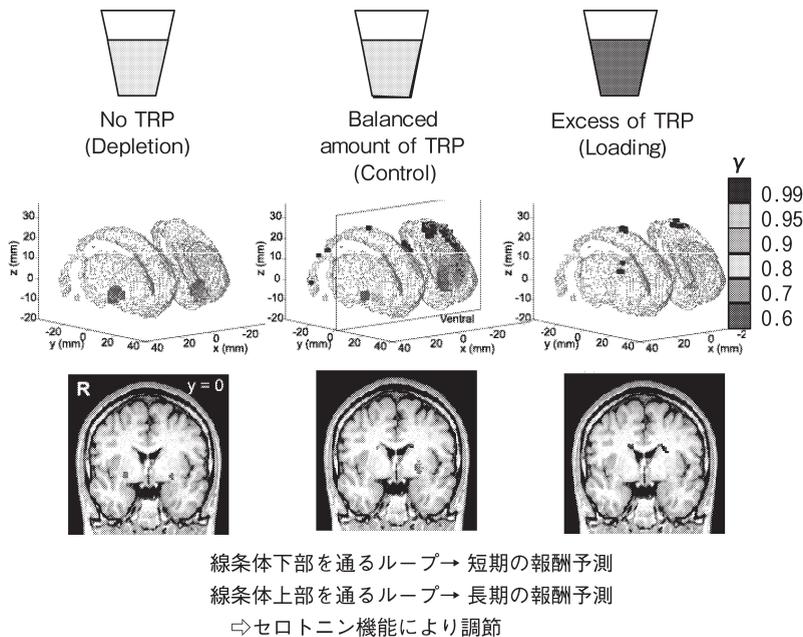


図3 各トリプトファン条件下で線条体の活性部位が異なる

結果とよく一致した。それに対し、セロトニンレベルが低い状態では、線条体の腹側部のみに短期的報酬予測に関わる活動が見られ、セロトニンレベルが高い状態では、線条体の背側部のみに長期的報酬予測に関わる活動が見られた(図3)。この結果は、脳の中に、異なる時間スケールで報酬予測を行う線条体を經由する並列ネットワークが存在し、セロトニンが線条体の活動を調節することでそれらのネットワークの活動を調節しているというメカニズムを示唆しており、セロトニンが報酬予測の時間スケールパラメータを調節するというわれわれの仮説を支持するものである。

上記のセロトニン操作を行ったfMRI実験では、行動データにおいてセロトニンによる条件の差がみられなかった。そこで、われわれは、上記の実験と同様のトリプトファン操作をした際の条件間の選択行動の差異を検討するため行動実験を行った<sup>14)</sup>。なお、本実験では、報酬はジュースではなく金銭的な報酬とした。その結果、小さな報酬の選択率において条件の効果がみられ、トリプ

トファン欠乏条件は、過剰条件よりも高い確率で小さい報酬を選択しており、対照条件よりも傾向差であるが、高い確率で小さい報酬を選択していることが示唆された。また、割引定数 $\gamma$ についても、トリプトファン欠乏条件の $\gamma$ は、過剰条件と対照条件にくらべて、低い結果となった。

これらの働きの脳内制御機構として銅谷らは、ラットを用いて同様の遅延報酬課題遂行中の背側縫線核によるセロトニン流量の増加<sup>9)</sup>やセロトニン神経系の発火の亢進<sup>7)</sup>が、遅延報酬を待つことに関連していることを明らかにしている。すなわち、ヒトにおいても遅延報酬に対する“割引”機能はセロトニンによって調節を受けるものと考えられた。

## 6. 精神疾患への応用の可能性

一方で、精神疾患のなかにもセロトニンが関与する疾患がある。うつ病は選択的セロトニン再取り込み阻害薬が治療上有効であり、脳機能画像研究においても、うつ病におけるセロトニン神経系

の機能異常が明らかとなってきた<sup>1,6)</sup>。

われわれは、うつ病の機能仮説として強化学習の理論の中で非常に重要な要素である報酬予測のメカニズムに障害をきたしていることを推測している。すなわち、うつ病の患者ではこの強化学習において将来の報酬を予測していく機能が障害されているため、「将来の報酬への見通し」が立たず、じっとしていること（行動抑制）や短絡的な行動（自殺、衝動行為）を最適行動として選択する。このような、うつ病の報酬予測障害仮説<sup>4,5)</sup>については、いまだうつ病症例での結果は得ておらず、今後のさらなる研究が必要である。

他方、短絡的な衝動行為に関連してアルコール依存症、病的賭博、オピオイド乱用、コカイン乱用、注意欠陥多動性障害（ADHD）、喫煙者など、衝動性を特徴とする疾患の患者では遅延による割引率の度合いが強い、すなわち同じ遅延の場合に価値をより小さく見積もることが示されている<sup>13)</sup>。さらに強迫性障害の病態に関するループと今回示した短期小報酬ループがほとんど重なっており、衝動的な強迫行為との関連が推定されている<sup>12)</sup>。

## 7. おわりに

今回、衝動性に関わる脳部位と脳内物質の知見を、数理モデルを軸につなぎ合わせ実験的に検証した結果を紹介した。昨今、多くの神経科学の研究者たちが、意志決定、情動、注意、意識などといった高次脳機能へ関心を寄せてきている。精神疾患を、これらの高次脳機能を理解するための障害モデルとしてとらえ、そのシステムを解明していくことは脳を知る上で非常に意義があると思われる。その上で、これらの脳機能の解明が精神疾患で苦しむ人々に福音となることを期待してやまない。

## 文 献

1) Bhagwagar, Z., Murthy, N., Selvaraj, S., et al.: 5-HTT binding in recovered depressed patients and healthy volunteers: a positron emission tomography study with [<sup>11</sup>C]DASB. *Am J Psychiatry*, 164; 1858-

1865, 2007

2) Bizot, J., Le Bihan, C., Puech, A.J., et al.: Serotonin and tolerance to delay of reward in rats. *Psychopharmacology (Berl)*, 146; 400-412, 1999

3) Cardinal, R.N., Pennicott, D.R., Sugathapala, C. L., et al.: Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science*, 292; 2499-2501, 2001

4) Doya, K.: Metalearning, neuromodulation, and emotion. *Affective Minds* (ed. by Hatano, G., et al.). Elsevier Science B. V., Amsterdam, p.101-104, 2000

5) Doya, K.: Modulators of decision making. *Nature Neurosci*, 11; 410-416, 2008

6) Meyer, J.H., Houle, S., Sagrati, S., et al.: Brain serotonin transporter binding potential measured with carbon 11-labeled DASB positron emission tomography: effects of major depressive episodes and severity of dysfunctional attitudes. *Arch Gen Psychiatry*, 61; 1271-1279, 2004

7) Miyazaki, K., Miyazaki, K.W., Doya, K.: Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. *J Neurosci*, 31; 469-479, 2011

8) Miyazaki, K.W., Miyazaki, K., Doya, K.: Activation of the central serotonergic system in response to delayed but not omitted rewards. *Eur J Neurosci*, 33; 153-160, 2011

9) Mobini, S., Chiang, T.J., Ho, M.Y., et al.: Effects of central 5-hydroxytryptamine depletion on sensitivity on delayed and probabilistic reinforcement. *Psychopharmacology (Berl)*, 152; 390-397, 2000

10) Mobini, S., Body, S., Ho, M.Y., et al.: Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology (Berl)*, 160; 290-298, 2002

11) O'Doherty, J.P., Hampton, A., Kim, H.: Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci*, 1104; 35-53, 2007

12) 岡本泰昌：強迫性障害の薬物療法とセロトニン・ドパミン仮説。精神経誌, 113; 36-44, 2011

13) 志々田一宏, 岡本泰昌：衝動性の神経科学的基盤。精神科治療学, 21; 807-815, 2006

14) Schweighofer, N., Bertin, M., Shishida, K., et al.: Low-serotonin levels increase delayed reward dis-

counting in humans. *J Neurosci*, 28 ; 4528-4532, 2008

15) Sutton, R.S., Barto, A.G. : Reinforcement Learning. MIT Press, Cambridge, 1998

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

17) Tanaka, S.C., Doya, K., Okada, G., et al. : Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci*, 7 ; 887-893, 2004

18) Tanaka, S.C., Schweighofer, N., Asahi, S., et

al. : Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum. *PLoS ONE*, 2 ; e1333, 2007

19) Thorndike, E.L. : Animal intelligence : An experimental study of the associative processes in animals. *Psychological Monographs*, 2 ; 1-109, 1898

20) Wogar, M.A., Bradshaw, C.M., Szabadi, E. : Effect of lesions of the ascending 5-hydroxytryptaminergic pathways on choice between delayed reinforcers. *Psychopharmacology (Berl)*, 111 ; 239-243, 1993

---

## Effects of Serotonin on Delay Discounting for Rewards

### —an Application for Understanding of Pathophysiology in Psychiatric Disorders—

Yasumasa OKAMOTO<sup>1)</sup>, Go OKADA<sup>1)</sup>, Kazuhiro SHISHIDA<sup>1)</sup>, Takuji FUKUMOTO<sup>1)</sup>,  
Akihiko MACHINO<sup>1)</sup>, Hidehisa YAMASHITA<sup>1)</sup>, Saori C TANAKA<sup>2)</sup>, Kenji DOYA<sup>3)</sup>, Shigeto YAMAWAKI<sup>1)</sup>

1) *Department of Psychiatry and Neurosciences, Hiroshima University*

2) *Institute of Economic Research, Osaka University*

3) *Neural Computational Unit, Okinawa Institute of Science and Technology*

In our daily life, we constantly make such choices between actions leading to rewards of various sizes after different delays. “Delay discounting” is a theoretical concept in which the “value” of reward R after delay. A steep rate of discounting results in impulsive choice, defined by an abnormally frequent choice of the more immediate reward. Our behavioral and neuroimaging results suggest that serotonin may adjust the rate of delayed reward discounting via the modulation of striatum in cortico-basal ganglia circuits in human. Our proposed role of serotonin may explain certain aspects of impulsivity in psychiatric disorders such as major depression, panic disorder or obsessive-compulsive disorder, that are known to effectively relieve symptoms by selective serotonin reuptake inhibitors. Future experiments using delayed reward paradigms could be designed to study impulsivity in these patients.

<Authors' abstract>

<Key words : psychiatric disorder, delayed discounting, reward, striatum, serotonin>