

**特集 2 計算論的精神医学の新展開：診断，病態解明，そして早期予測への革新的アプローチ**

## 1. ギャンブル依存と強化学習

鈴木 真介\*

**抄録：**わが国においてもカジノ解禁を見据えた動きが進むなかで、ギャンブル依存（ギャンブル障害）は重要な社会問題として注目されている。本稿では、ギャンブル障害患者と健常者における意思決定プロセスの差異を、強化学習モデルを用いて検証する試みを紹介する。まず、「強化学習と意思決定に関する計算論的神経科学」の研究を概観する。次に、「悪い結果をもたらすにもかかわらずギャンブルをやめられない」というギャンブル障害の症状と密接にかかわる「行動の柔軟性の欠如」が、強化学習の枠組みでどのように表現できるかについて議論する。最後に、筆者らが実施した計算論的精神医学の研究を紹介し、強化学習を用いた行動データのモデリングと機能的磁気共鳴画像法（fMRI）を組み合わせて、「ギャンブル障害患者における行動の柔軟性の欠如」の神経基盤を検証した結果について述べる。

日本生物学的精神医学会誌 36 (1) : 27-30, 2025

**Key words :** computational psychiatry, gambling disorder, reinforcement learning, functional MRI, computational modeling

ギャンブル障害（ギャンブル依存）は、「臨床的に意味のある機能障害または苦痛を引き起こすに至る持続的かつ反復性の問題賭博行為」と定義されている<sup>1)</sup>。わが国においては、過去1年にギャンブル等依存が疑われる者の割合は2.2%と推計されており（令和2年度依存症に関する調査研究事業「ギャンブル障害およびギャンブル関連問題の実態調査」報告書）、その問題の深刻さが伺える。また、2018年にはカジノ解禁を見据えて「ギャンブル等依存症対策基本法」が施行され、ギャンブル依存が社会的な課題として重要視されている。本稿では、機能的磁気共鳴画像法（functional magnetic resonance imaging : fMRI）を用いた脳機能計測実験と、強化学習を用いた行動データのモデリングを組み合わせ、ギャンブル障害患者と健常者における意思決定プロセスの差異を検証した筆者らの計算論的精神医学研究<sup>8)</sup>について紹介する。

### 1. 強化学習と報酬に基づく意思決定

初めてわが国を訪れた旅行者がホテルの近くにあ

るAとBの2つのコンビニエンスストアを見つけたとする。旅行者は最初、どちらのコンビニエンスストアが好みに合っているかの情報がないため、ランダムに選んで買い物をする。しかし、何度か訪れるうちに、どちらの店が自分の好みに合っているかを学習し、より適切な選択をできるようになる。このような意思決定のプロセスは、動物の採餌行動からヒトの社会経済行動に至るまで、さまざまな場面で広くみられる。

上記のような意思決定は、機械学習アルゴリズムの一種である「強化学習」で記述することができる<sup>5,7)</sup>。強化学習の枠組みでは、エージェント（意思決定者）は各選択肢の価値を計算し、価値の高い選択肢をより高い確率で選ぶことが想定されている。たとえば、先の例では、コンビニエンスストアAの価値 $Q(A)$ とBの価値 $Q(B)$ が脳内で計算され、それに基づいて選択が行われる。選択の後、その結果得られた「報酬」に応じて当該選択肢の価値が更新される。具体的には、もしAを選択して美味しい（＝高報酬）食品を購入できればAの価値が増加し、逆に購入した食品が美味しくなかった

Reinforcement learning in gambling disorder

\*一橋大学ソーシャル・データサイエンス研究科（〒186-8601 東京都国立市中2-1）Shinsuke Suzuki : Faculty of Social Data Science, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan

【鈴木 真介 E-mail : shinsuke.szk@gmail.com】

(=低報酬) 場合, A の価値は減少する。この更新プロセスは, 「報酬予測誤差」に基づいて行われ, 次の式で表される。

$$Q_{t+1}(A) = Q_t(A) + \alpha\delta. \quad (1)$$

ここで,  $\delta = R_t - Q_t(A)$  は報酬予測誤差を意味し, 得られた報酬  $R_t$  と更新前の価値  $Q_t(A)$  の差を示す。つまり, 得られた報酬が予想を上回れば価値は増加し, 予想を下回れば価値は減少する (予想どおりであれば変化しない)。また,  $0 \leq \alpha \leq 1$  は学習率とよばれ, 学習 (=価値の更新) の速さを決定する。

強化学習モデルは, ヒトや動物の行動だけでなく, その背後にある脳の活動を説明できることが知られている。動物実験, 特にサルや齧歯類を対象とした電気生理学的研究では, 中脳ドーパミン神経細胞の活動が報酬予測誤差を表現していることが示されている<sup>6)</sup>。さらに, ヒトを対象とした脳機能計測実験では, ドーパミン神経細胞の主要な投射先である線条体, 前頭前野内側部, 島皮質などの領域における脳活動 (fMRI 信号) が, 報酬予測誤差と相関することが報告されている<sup>2)</sup>。これらの知見は, ヒトの脳, 特にドーパミン系が, 強化学習という機械学習アルゴリズムを実装しており, 日常の学習や意思決定に寄与していることを示唆している。

## 2. ギャンブル障害と強化学習

「悪い結果をもたらすにもかかわらずギャンブルを止められない」というギャンブル障害の症状は, 「行動の柔軟性の欠如」と密接に関連していると考えられている。では, この柔軟性の欠如は強化学習の枠組みでどのように表現されるのか? 2つのアプローチが提案されている<sup>4)</sup>。1つ目は, 強化学習に「意思決定の慣性 (decision inertia)」を導入するアプローチである。このアプローチでは, 意思決定の際に「前回選ばれた選択肢の価値」に正の値が加算される。これにより, 前回と同じ選択肢が再び選ばれる可能性が高まり, その結果として行動の柔軟性が低下することが説明される。

2つ目のアプローチは「学習率の非対称性」の導入である。これは, 報酬予測誤差の正負に応じて異なる学習率が適用されるというものである。数式としては次のように表現される。

$$\alpha = \begin{cases} \alpha_{(+)} & \text{if } \delta \geq 0 \\ \alpha_{(-)} & \text{if } \delta < 0 \end{cases} \quad (2)$$

つまり, 正の報酬予測誤差に対しては学習率  $\alpha_{(+)}$ ,

負の報酬予測誤差に対しては学習率  $\alpha_{(-)}$  が適用される。非対称な学習率のもとでは, 正の報酬予測誤差に高い学習率, 負の報酬予測誤差に低い学習率が適用されるとき, 前回選ばれた選択肢が再度選ばれる可能性が増加し柔軟性の欠如が表現され得る。正の報酬予測誤差に対する高い学習率のもとでは, 直近の行動の「よい結果」が大きく評価され, その行動が繰り返される可能性が増加する。一方, 負の報酬予測誤差に対する低い学習率のもとでは, 直近の行動の「悪い結果」が十分に考慮されず, 行動の切り替えが行われにくい。このようなメカニズムによって, 前回と同じ選択肢が再び選ばれる可能性が高まり, 柔軟性の欠如が説明される。

筆者らの研究では, ギャンブル障害患者と健常者の間で「意思決定の慣性」や「学習率」に差があるかを検証した<sup>8)</sup>。また, 同時に fMRI を用いて脳活動を測定し, その神経基盤についても調べた。実験には, 年齢, 性別, 知能指数がおおむね一致した 17 名のギャンブル障害患者と 34 名の健常者が参加した。各被験者は MRI 装置内で, ギャンブルを模した「道具的条件づけ」とよばれる意思決定課題を行った。この課題では, 被験者は 2 つの図形から 1 つを選ぶ。各選択の後には, 選んだ図形に応じて確率的に報酬が与えられる。被験者はこのプロセスを繰り返すことで, 2 つの図形の報酬確率を学習し, より高い報酬を得る可能性のある図形を選ぶようになる。

この意思決定課題において, 被験者はどのように行動をしていたのか? ギャンブル障害患者と健常者の間に行動様式の違いはあったのか? 筆者らはまず, 意思決定の慣性や非対称な学習率などさまざまなパラメータを含む複数の強化学習モデルを構築し, 実際の被験者の行動パターンと照合した。その結果, ギャンブル障害患者と健常者の行動をもっともよく説明するモデルは「非対称な学習率をもつ強化学習」であることがわかった。

次に, 上記のモデルを用いて各被験者の学習率を推定し, ギャンブル障害患者は健常者と比べて「正の報酬予測誤差に対する学習率が高く, 負の報酬予測誤差に対する学習率が低い」という特徴を見いだした (図 1)。つまり, ギャンブル障害患者は予想以上の報酬 (=ギャンブルにおける勝利) から過大に学習し, 一方で予想以下の報酬 (=ギャンブルにおける敗北) からは過小に学習する傾向があるということである。この結果は, ギャンブル障害にみられる「行動の柔軟性の欠如」の一因が非対称な学習率にある可能性を示唆している。すなわち, ギャン

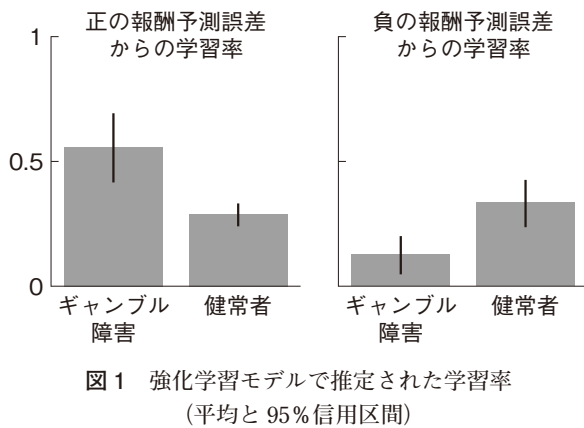


図1 強化学習モデルで推定された学習率 (平均と95%信用区間)

ブル障害患者は勝利の経験を過大に評価し、その行動を繰り返す一方、敗北の経験は十分に学習されず、新たな選択肢を試すことが少なくなる。このような行動の固定化が、ギャンブル障害の症状である「賭博をやめられない」状況に作用していると考えられる。

ギャンブル障害患者と健常者の学習様式の違いは、どのような脳の情報処理を反映しているのか？筆者らは、脳活動 (fMRI 信号) データの解析を行い、これを検証した。その結果、ギャンブル障害患者においては健常者と比べて、「正の報酬予測誤差に対する島皮質の脳活動の増加分が大きい (図2)」ことが明らかになった。これは、ギャンブル障害患者の島皮質が正の報酬予測誤差に対してより敏感に反応していることを示唆する。一方で、負の報酬予測誤差に対する脳活動には両群間で顕著な差異はみられなかった。これらの結果は、ギャンブル障害患者は健常者に比べて、島皮質の脳活動が正の報酬予測誤差に対してより鋭敏に反応することで、望ましい結果から過大に学習し、それが行動の柔軟性の欠如につながっている可能性を示唆している。

本稿では、筆者らの fMRI を用いた脳機能計測実験と強化学習を用いた行動データのモデリングを組み合わせた研究を紹介した<sup>8)</sup>。ギャンブル障害患者における意思決定のメカニズムについては、探索行動やリスク選好に関する研究なども行われている<sup>3, 9)</sup>。今後、計算論的精神医学のさらなる進展により、ギャンブル障害の背後にある脳の情報処理メカニズムが解明され、より効果的な治療法の開発につながることが期待される。

本論文に記載した筆者らの研究に関して、すべて倫理的配慮を行っている。開示すべき利益相反は存在しない。

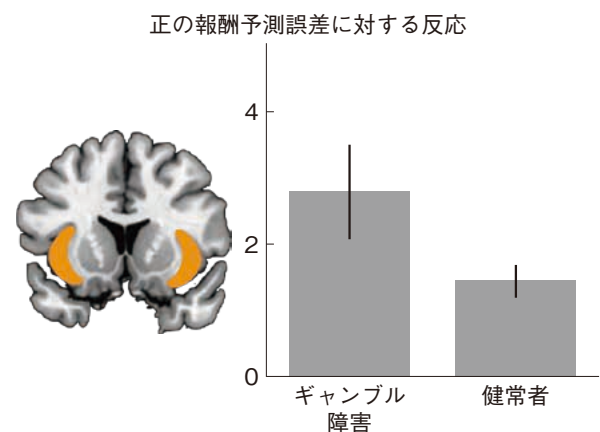


図2 正の報酬予測誤差に対する島皮質の反応 (fMRI 一般線形モデル解析の回帰係数: 平均と標準誤差)

## 文 献

- 1) American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, fifth edition (DSM-5). American Psychiatric Association, Washington, D. C.
- 2) Corlett PR, Mollick JA and Kober H (2022) Meta-analysis of human prediction error for incentives, perception, cognition, and action. *Neuropsychopharmacology*, 47 (7) : 1339-1349.
- 3) Fujimoto A, Tsurumi K, Kawada R, et al (2017) Deficit of state-dependent risk attitude modulation in gambling disorder. *Transl Psychiatry*, 7 (4) : e1085.
- 4) Katahira K (2018) The statistical structures of reinforcement learning with asymmetric value updates. *J Math Psychol*, 87 : 31-45.
- 5) Rangel A, Camerer C, and Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci*, 9 (7) : 545-556.
- 6) Schultz W, Dayan P and Montague PR (1997) A neural substrate of prediction and reward. *Science*, 275 (5306) : 1593-1599.
- 7) Sutton RS and Barto AG (1998) Reinforcement learning : an introduction. MIT Press, Cambridge.
- 8) Suzuki S, Zhang X, Dezfouli A, et al (2023) Individuals with problem gambling and obsessive-compulsive disorder learn through distinct reinforcement mechanisms. *PLoS Biol*, 21 (3) : e3002031.
- 9) Wiehler A, Chakroun K and Peters J (2021) Attenuated directed exploration during reinforcement learning in gambling disorder. *J Neurosci*, 41 (11) : 2512-2522.

■ ABSTRACT

---

**Reinforcement learning in gambling disorder**

Shinsuke Suzuki

*Faculty of Social Data Science, Hitotsubashi University*

With the movements toward the legalization of casinos in Japan, gambling disorder is drawing attention as a significant social issue. Here, this article reviews an attempt to examine the differences in decision-making processes between individuals with gambling disorder and healthy controls through the application of reinforcement learning models to behavioral and neuroimaging data. First, I provide an overview of computational neuroscience research, particularly focusing on decision-making and reinforcement learning. The concept of “lack of behavioral flexibility,” which is closely linked to a core symptom of gambling disorder (i. e., difficulty in stopping gambling despite adverse outcomes), is then discussed within the context of reinforcement learning. Finally, I present our recent study in computational psychiatry, which explores the neural basis of behavioral inflexibility in individuals with gambling disorder by combining reinforcement learning and functional magnetic resonance imaging (fMRI). No potential conflicts of interest were disclosed.

(Japanese Journal of Biological Psychiatry 36 (1) : 27–30, 2025)

---