

報酬と意思決定の実時間脳内機構の研究 (052107012)

Brain mechanism of real-time processing for reward and decision making

研究代表者

鮫島和行 玉川大学脳科学研究所

Kazuyuki Samejima, Brain Science Institute, Tamagawa University

研究期間 平成 17 年度～平成 19 年度

本研究開発の概要

人間および動物は、不確実で変動するような環境において適切な意思決定を行い、食べ物や飲み物などの生存に不可欠な報酬を得ることによって生き抜いている。本研究では、強化学習とよばれる計算論的モデルとサルとヒトを用いた行動中の脳活動計測などの実験的手法を組み合わせる報酬に基づく意思決定の脳内機構を研究した。特に、サルの線条体において特定の行動における報酬予測の表現が存在することを示した。また、ヒトの fMRI 研究で被験者毎に異なる長期報酬予測を行動から求め、それと相関する活動が線条体に存在することを示した。このことは大脳皮質と大脳基底核で結ばれるループ回路が報酬に基づく意思決定において、報酬の予測と決定に重要な役割を果たすことを示唆する。

Abstract

Animals including humans can decide appropriate actions leading to get future reward, such as food or water, even in dynamic and probabilistic environment. In this study, I developed a new method combining a theoretical and experimental approach to study reward based decision-making in animals and humans. By applying reinforcement learning (RL) theory to predict monkey and human behavior and brain activities, we show that the cortico-basal ganglia network play an important role for reward-based decision making by predicting and comparing action-specific reward expectations for an action candidate.

1. まえがき

我々が意思決定を行うとき、常に選択ののちの報酬等の結果を考慮して最善の選択をする。報酬に基づく意思決定と、その学習を不確実で変動するような環境で行う、ということは我々の脳の生存に対する重要な機能である。このような報酬に基づく意思決定の学習は「強化学習」と呼ばれる問題設定として定式化され、その解法のいくつかが脳の計算論的モデルとして提案されてきた。その 1 つに行動選択肢ごとに報酬を予測し、それを比較することで意思決定を行う「行動価値を用いる強化学習法」が大脳皮質-大脳基底核回路で計算される、という仮説を提案し、それを検証するための生理学実験及び、fMRI を用いたヒトの非侵襲脳活動計測による実験的検証を行った。

2. 研究内容及び成果

2.1 線条体における行動価値表現と強化学習

強化学習モデルは、1)行動を取った後の結果を事前に予測する(行動価値)、2)それらを比較してより良い行動を選択する(行動選択)、3)選択した後の結果(報酬)と予測との差分(報酬予測誤差)によって行動価値を更新する、という 3 つのプロセスによって報酬に基づく意思決定とその学習を説明する。

強化学習モデルが大脳皮質-大脳基底核ループ回路に再現されるという仮説が提案されている。この仮説を検証するために、2 つの行動選択肢から 1 つを選んで実行する課題をサルに訓練し、課題遂行中の線条体の神経活動を記録した。各行動後に得られる報酬は確率的に与えられ、その確率は一定試行毎に変動させた。その結果、サルは自ら選んだ行動と報酬に基づいて、より多くの報酬が得られるように選択行動を変化させた(図 1)。サルが選択した行動と報酬の履歴を強化学習モデルに適用すると、サルが次に行う行動の選択確率を予測できた。このサルの行動をシミュ

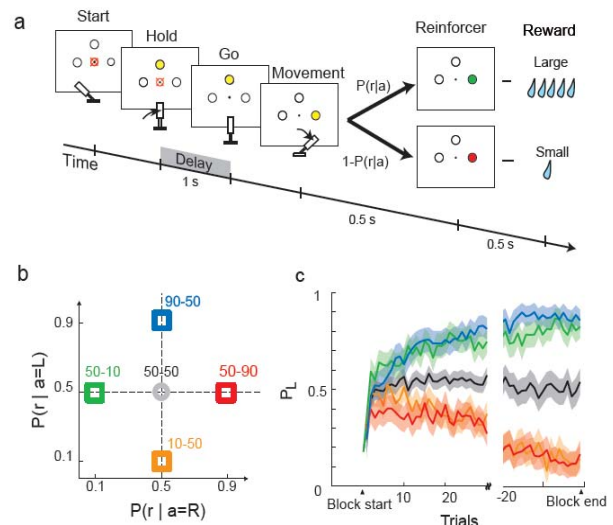


図 1. 報酬に基づく意思決定課題とサルの行動パフォーマンス。 a)タスクの 1 試行の流れ b) 左右の報酬確率 $P(r|a=L)$, $P(r|a=R)$ の組み合わせによる 5 種類のブロック(青 90-50、緑 50-10、黄色 10-50、赤 50-90:各数字は左・右の報酬確率を表す c) 行動選択の平均確率(実線)と 95%信頼区間(帯)、色はブロックの種類を表し b) と同様 (Samejima et al 2005 より)

ュレートしたモデルは、直接には観測できない毎試行変動する行動価値と、サルの線条体神経活動の比較を可能にする。比較の結果、サルの行動選択直前の線条体投射細胞の神経活動では、行動選択(ΔQ -type)は少なく、行動ごとの報酬予測、すなわち行動価値(Q_L/Q_R type)に相関する細胞が多いことを見出した(図 2)。

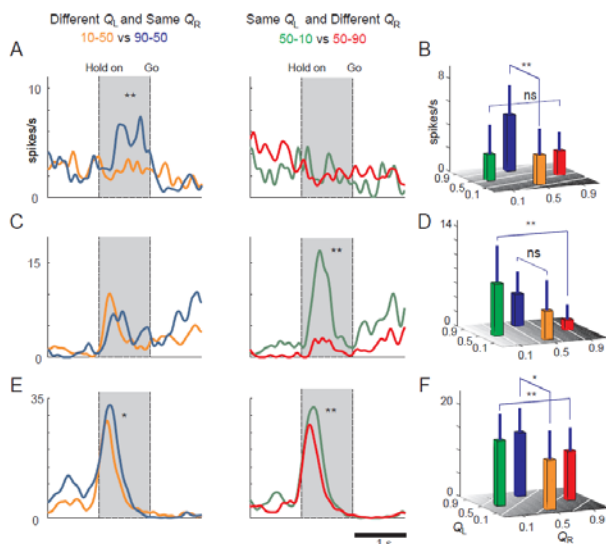


図 2. 線条体神経活動の代表 3 例 (A) 前部線条体から記録された左の行動価値を表現する神経活動(QI-type)。左は 10-50, 90-50 ブロックの平均発火頻度 (left panel) 右は 50-10, 50-90 ブロック。(B) 4 つの異なるブロックでの運動直前 (A での灰色区間) の発火強度 (spikes/sec) を 3 次元棒グラフで表示。床面の白黒の明るさと等高線 (白線) は左右の行動価値による神経活動の回帰平面を表す。(C, D) 前部被殻から記録された右の行動価値タイプ(QR-type) の神経活動 (E, F) 左右の行動価値の差分、または行動そのものに関する神経活動 (ΔQ & m-type) (Samejima et al 2005 より)

これらの結果は、1)線条体において行動価値が表現され、2)それが大脳基底核の下流に送られることによって行動選択が起こり、3)報酬予測誤差を表現する中脳ドーパミン細胞からの線条体への投射によって行動価値が学習される、という大脳基底核-強化学習仮説を支持している。

2.2 認知的意思決定時の脳活動計測

大脳基底核が価値関数を用いた強化学習を実現しているということを示してきた。これら線条体における価値表現は、大脳皮質へと戻るループ回路を通じて 1 つの行動を選択するために使われるであろうか？決定が生じるための価値の比較と選択は脳のどの回路において行われているだろうか？この疑問を解決するために、今回運動そのものを誘発する脳活動や、運動を準備する脳活動と、報酬の比較と選択の脳活動を時間的に分離する新しいパラダイムを提案し、fMRI を用いて意思決定課題中の脳活動を計測した。

図 3 に示すような 4 種類の色と 4 種類の形を持つ 16 種類のターゲット刺激からランダムに 2 つを選んで画面の左右に呈示(Target)し、被験者は右手に持つ左右のボタン押しによって選択を行う。その結果、あるブロックでは色ごとに、またあるブロックでは形ごとに、あらかじめ決められた金額の報酬を得る。Target 呈示の 2~6 秒前に、提示される Target の色、または形を事前呈示 (Cue) する。事前呈示される Cue が報酬に関係のある情報である場合には、被験者は Cue の呈示時点で選択すべき Target の色、または形を知り認知的に意思決定することができる。しかし、報酬に関係のない情報である場合には、Target が提示されるまで選択すべき対象はわからないため、意思決定は不可能となる。

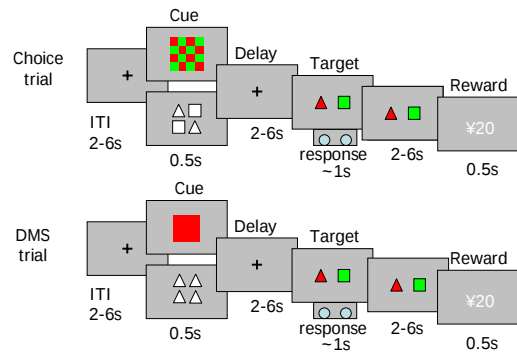


図 3. 認知的意思決定課題の 1 試行の流れ

12 人の被験者を用いて fMRI により BOLD 信号を計測し、信号が Cue 刺激で一過性に活動する領域を Event related analysis によって求めたところ、選択する色または形の報酬量予測の回帰係数は腹側線条体領域、特に側座核と腹側被殻において、有意な正の相関が認められた。このことは、認知的意思決定に伴う報酬予測が線条体、特に腹側線条体で行われていること示唆している。さらに、認知的意思決定に特異的な脳活動を調べるために、認知的に意思決定できる Cue と認知的に意思決定できない Cue との差分の活動を尾側帯状回運動皮質に見出した。このことは、大脳基底核によって計算された価値関数を比較することによって望ましいターゲットを選択する過程に帯状皮質運動皮質が関係することを示唆する。

3. むすび

本研究課題において、大脳基底核における強化学習という計算論的モデルの検証を行い、線条体が行動価値の情報表現を持つことを明らかにし、また前頭葉の帯状回皮質が価値の比較と選択に重要であることを示唆する非侵襲計測実験を行った。同時に、ヒトでの課題と同様の認知的意思決定課題をサルに訓練したが、大脳基底核の神経活動を報告するにはいたらなかった、今後はこの研究課題で開始した研究を継続し、行動決定学習における大脳基底核と前頭葉の情報表現と処理様式から、計算論を明らかにしていきたい。

【誌上発表リスト】

- [1] Samejima K, Ueda Y, Doya K, and Kimura M, Representation of Action-specific-reward value in the striatum, *Science*, **310** (5752), 1337-1339 (2005)
- [2] Samejima, K. & Doya, K., Multiple representations of belief states and action values in corticobasal ganglia loops., *Ann N Y Acad Sci.*, **1104**, 213-228 (2007)
- [3] Kawato, M. & Samejima, K., Efficient reinforcement learning: computational theories neuroscience and robotics., *Curr Opin Neurobiol.* **17** (2), 205-212 (2007)

【受賞リスト】

- [1] 鮫島和行、上田康雅、銅谷賢治、木村實、日本神経回路学会論文賞、2006

【報道発表リスト】

- [1] ATR、京都府立医科大学 同時プレスリリース “自由意志による行動選択を実現する脳の計算過程～選択肢のメリットを評価する神経細胞を発見～” 2005 年 11 月 25 日